Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2332533 A1 2001/08/18

(21) 2 332 533

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(22) Date de dépôt/Filing Date: 2001/02/16

(41) Mise à la disp. pub./Open to Public Insp.: 2001/08/18

(30) Priorité/Priority: 2000/02/18 (60/184,152) US

(51) Cl.Int./Int Cl. C12N 15/57, A61K 38/48, C12N 9/64, C12N 5/00, C07K 16/40, C12N 5/10, C12Q 1/00, C12P 21/00

(71) Demandeur/Applicant:
F. HOFFMANN-LA ROCHE AG, CH

(72) Inventeurs/Inventors:
ZUO, FENGRONG, US;
KLONOWSKI, PAUL, US;
HELLER, RENU ANAND, US

(74) Agent: BORDEN LADNER GERVAIS LLP

(54) Titre : NOUVELLES METALLOPROTEASES COMPORTANT DES DOMAINES DE NATURE THROMBOSPONDINE ET COMPOSITIONS D'ACIDES NUCLEIQUES CODANT CES COMPOSES

(54) Title: NOVEL METALLOPROTEASES HAVING THROMBOSPONDIN DOMAINS AND NUCLEIC ACID COMPOSITIONS ENCODING THE SAME

(57) Abrégé/Abstract:
Novel metalloproteases having thrombospondin domain(s) (MPTS proteins) and polypeptides related thereto, as well as nucleic acid compositions encoding the same, are provided. The subject polypeptide and nucleic acid compositions find use in a variety of applications, including diagnostic applications, therapeutic agent screening applications, as well as therapeutic applications for a variety of different conditions. Also provided are methods of treating disease conditions associated with aggrecanase activity, e.g. conditions characterized by the presence of aggrecan cleavage products, such as rheumatoid- and osteo-arthritis.

Abstract

Novel metalloproteases having thrombospondin domain(s) (MPTS proteins) and polypeptides related thereto, as well as nucleic acid compositions encoding the same, are provided. The subject polypeptide and nucleic acid compositions find use in a variety of applications, including diagnostic applications, therapeutic agent screening applications, as well as therapeutic applications for a variety of different conditions. Also provided are methods of treating disease conditions associated with aggrecanase activity, e.g. conditions characterized by the presence of aggrecan cleavage products, such as rheumatoid- and osteo-arthritis.

The field of the invention is proteases, particularly metalloproteases with thrombospondin domains.

Cartilage matrix structure as dry weight of the tissue is made up of 70% collagen and 20-30% proteoglycans. The proteoglycan component confers mechanical flexibility to load bearing tissues and imparts viscoelastic properties to cartilage. Its loss leads to rapid structural damage as is seen most frequently in arthritic joint diseases and joint injury.

Aggrecan is a major cartilage proteoglycan. Aggrecan is a large protein of 210 kDa and has three globular domains: G1, G2, and G3. The G1 and G2 domains of the protein are closer to the amino terminus of the protein and their intervening interglobular domain has sites that are proteolytically sensitive. The region between G2 and G3 is heavily glycosylated and connected to oligosaccharides and glycosaminoglycans (GAGs) to form the mature proteoglycan. In arthritic cartilage, core protein fragments of 55 kDa are observed and believed to be the result of cleavage of the core protein in the G1 and G2 interglobular domain between asparagine 341 and phenylalanine 342. This cleavage can be made by many matrix metalloproteinases e.g. MMP-1, -2, -3, -7, -8, -9, and -13. In addition, 60 kDa aggrecan fragments with a -COOH terminus of glutamic acid are also identified and are indicative of a cleavage site between glutamic acid 373 and alanine 374. Matrix metalloproteinase are unable to cleave at this site. The unique endopeptidase activity responsible for this cleavage has been termed "aggrecanase."

The G1 domain of the core protein forms a stable ternary complex by binding to hyaluronic acid and link proteins in the matrix. Any enzymatic cleavage in this region destabilizes the cartilage matrix structure, leads to the loss of the major proteoglycan aggrecan and exposes type II collagen to collagenases, causing cartilage loss and the consequent development of joint disease. Since a variety of anti-arthritic drugs do not

As such, aggrecanase is considered to be an important drug target for arthritis. Aggrecan fragments released into the synovial fluid are the primary detectable events in the development of rheumatoid- and osteo- arthritis. Search for this protease has been intense. Despite these intense discovery efforts, identification of human aggrecanase has remained elusive.

As such, there is much interest in the identification of human aggrecanase, as well as the gene encoding this activity.

The following references are directed to this field. U.S. Patents 5,872,209 and 5,427,954, and WO 99/09000; WO 98/55643; WO 98/51665; and WO 97/18207.

Other references include: Abbasdale, "Cloning and characterization of ADAMTS11, an aggrecanase from the ADAMTS family," J. Biol. Chem. (Aug. 1999) 274: 23443-50; Arner et al., "Generation and Characterization of Aggrecanase. A soluble, cartilage-derived aggrecan-degrading activity," J Biol Chem (1999 Mar 5) 274(10):6594-6601; Arner et al., "Cytokine-induced cartilage proteoglycan degradation is mediated by aggrecanase," Osteoarthritis Cartilage (1998 May) 6(3):214-28; Billington et al., "An aggrecan-degrading activity associated with chondrocyte membranes," Biochem J (1998 Nov 15) 336 (Pt 1):207-12; Buttner et al., "Membrane type 1 matrix metalloproteinase (MT1-MMP) cleaves the recombinant aggrecan substrate rAgg1mut at the 'aggrecanase' and the MMP sites. Characterization of MT1 MMP catabolic activities on the interglobular domain of aggrecan," Biochem J (1998 Jul 1)333 (Pt 1):159-65; Flannery et al., "Expression of ADAMTS homologues in articular cartilage," Biochem. Biophys. Res. Commun. (July 1999) 260:318-22; Hurskainen et al., "ADAM-TS5, ADAM-TS6, and ADAM-TS7, Novel members of a New Family of Zinc Metalloproteases," J. Biol. Chem. (Sept. 1999) 274: 25555-25563Hughes et al., "Differential expression of aggrecanase and matrix metalloproteinase activity in chondrocytes isolated from bovine and porcine articular cartilage," J Biol Chem (1998 Nov 13) 273(46):30576-82; Ilic et al., "Characterization of aggrecan retained and lost from the extracellular matrix of articular cartilage. Involvement of carboxyl-terminal processing in the catabolism of aggrecan," J Biol Chem (1998 Jul 10) 273(28):17451-8; Kuno et al., "ADAMTS-1 is an active metalloproteinase associated with the extracellular

matrix," J. Biol. Chem. (June 1999) 274:18821-6; Kuno et al., "ADAMTS-1 protein anchors at the extracellular matrix through the thrombospondin type I motifs and its spacing region," J. Biol. Chem. (May 1998) 273:13912-7; Kuno et al., "The exon/intron organization and chromosomal mapping of the mouse ADAMTS-1 gene encoding an ADAM family protein with TSP motifs," Genomics (Dec. 1997) 46:466-71; Kuno et al., "Molecular cloning of a gene encoding a new type of metalloproteinase-disintegrin family protein with thombospondin motifs as an inflammation associated gene," J. Biol. Chem. (Jan. 1997) 272: 556-62; Sandy et al., "Chondrocyte-mediated catabolism of aggrecan: aggrecanase-dependent cleavage induced by interleukin-1 or retinoic acid can be inhibited by glucosamine," Biochem J (1998 Oct 1) 335 (Pt 1):59-66; Tang & Hong, "ADAMTS: a novel family of proteases with ADAM protease domain and thrombospondin 1 repeats," FEBS Lett. (Feb. 1999) 445:223-5; Tortorella et al., Purification and cloning of aggrecanase-1: a member of the ADAMTS family of proteins," Science (June 1999) 284:1664-6; Vankemmelbeke et al., "Coincubation of bovine synovial or capsular tissue with cartilage generates a soluble 'Aggrecanase' activity," Biochem Biophys Res Commun (1999 Feb 24) 255(3):686-91; and Vasquez et al., "METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new family of proteins with angio-inhibitory activity," J. Biol. Chem. (Aug. 1999) 274:23349-57.

The present invention is directedo to novel metalloproteases having thrombospondin domain(s) (MPTS proteins) and polypeptides related thereto, as well as nucleic acid compositions encoding the same, are provided. The subject polypeptide and nucleic acid compositions find use in a variety of applications, including diagnostic applications, therapeutic agent screening applications, as well as therapeutic applications for a variety of different conditions. Also provided are methods of treating disease conditions associated with aggrecanase activity, e.g. conditions characterized by the presence of aggrecan cleavage products, such as rheumatoid- and osteo-arthritis.

MPTS protein of the subject invention. Figure 1B provides the amino acid sequence

of MPTS -15. Figure 1C provides an alignment of the amino acid sequence of the subject MPTS-15 with the amino acid sequence of ADAMTS-6, a sequence disclosed in Hurskainen et al., J. Biol. Chem. (Sept. 1999) 274: 25555-25563.

Figure 2A provides the sequence of a nucleic acid that encodes MPTS-10, an MPTS protein of the subject invention. Figure 2B provides the amino acid sequence of MPTS-10.

Figure 3A provides the sequence of a nucleic acid that encodes MPTS-19, an MPTS protein of the subject invention. Figure 3B provides the amino acid sequence of MPTS-19.

Figure 4A provides the sequence of a nucleic acid that encodes MPTS-20, an MPTS protein of the subject invention. Figure 4B provides the amino acid sequence of MPTS-20.

Novel MPTS proteins and polypeptides related thereto, as well as nucleic acid compositions encoding the same, are provided. The subject polypeptide and/or nucleic acid compositions find use in a variety of different applications, including research, diagnostic, and therapeutic agent screening/discovery/ preparation applications. Also provided are methods of treating disease conditions associated with MPTS, including aggrecanase, function, e.g. diseases characterized by the presence of aggrecan cleavage products, such as rheumatoid- and osteo-arthritis.

Novel metalloproteases having thrombospondin domain(s) (also known as MPTS proteins, ADAMTS proteins or aggrecanase proteins), as well as polypeptide compositions related thereto, are provided. The term polypeptide composition as used herein refers to both the full length protein, as well as portions or fragments thereof. Also included in this term are variations of the naturally occurring human protein, where such variations are homologous or substantially similar to the naturally occurring protein, as described in greater detail below. In the following description of the subject invention, the term "MPTS" is used to refer not only to the specific human MPTS proteins disclosed herein (i.e. MPTS-10; MPTS-15; MPTS-19 and MPTS-20), but also to homologs thereof expressed in non-human species, e.g. murine, rat and other mammalian species.

Specific human MPTS proteins of interest are MPTS-15, MPTS-10, MPTS-19 and MPTS-20. MPTS-15 has an amino acid sequence as shown in Fig. 1B and identified as SEQ ID NO:01. MPTS-10 has an amino acid sequence as shown in Fig. 2B and identified as SEQ ID NO:03. MPTS-19 has an amino acid sequence as shown in Fig. 3B and identified as SEQ ID NO:05. MPTS-20 has an amino acid sequence as shown in Fig. 4B and identified as SEQ ID NO:07. The subject MPTS proteins have a molecular weight based on their amino acid sequence of at least about 90 kDal, where the molecular weight based on the amino acid sequence may be substantially higher in certain embodiments. The true molecular weight of the subject MPTS proteins may vary due to glycosylation and/or other postranslational modifications.

Also provided by the subject invention are MPTS polypeptide compositions. The term polypeptide composition as used herein refers to both the full length proteins as well as portions or fragments thereof. Also included in this term are variations of the naturally occurring proteins, where such variations are homologous or substantially similar to the naturally occurring protein, be the naturally occurring protein the human protein, mouse protein, or protein from some other species which naturally expresses an MPTS protein, usually a mammalian species. A candidate homologous protein is substantially similar to an MPTS protein of the subject invention, and therefore is an MPTS protein of the subject invention, if the candidate protein has a sequence that has at least about 35%, usually at least about 45% and more usually at least about 60% sequence identity with an MPTS protein, as determined using MegAlign, DNAstar (1998) clustal algorithm as described in D.G. Higgins and P.M. Sharp, "Fast and Sensitive multiple Sequence Alignments on a Microcomputer," (1989) CABIOS, 5: 151-153. (Parameters used are ktuple 1, gpa penalty 3, window, 5 and diagonals saved 5). In the following description of the subject invention, the term "MPTS-protein" is used to refer not only to the human MPTS proteins, but also to homologs thereof expressed in non-human species, e.g. murine, rat and other mammalian species.

[illegible] usually at least about 65% and more usually at least about 70%. In many preferred

embodiments, the sequence identity is at least about 90%, usually at least about 95% and more usually at least about 99% over the entire length of the protein.

In many embodiments, the proteins of the subject invention are enzymes, particularly proteinases and more particularly a metalloproteinases. The subject proteins of this embodiment are characterized by having aggrecanase activity. As such, the subject proteins are capable of cleaving aggrecan in an interglobular domain, particularly between the G1 and G2 domains, and more particularly at the Glu^{373}-Ala^{374} bond of human aggrecan, to produce a cleavage product having an N-terminal sequence of ARGSVIL.

In addition to the proteins described above, homologs or proteins (or fragments thereof) from other species, i.e. other animal or plant species, are also provided, where such homologs or proteins may be from a variety of different types of species, usually mammals, e.g. rodents, such as mice, rats; domestic animals, e.g. horse, cow, dog, cat; and humans. By homolog is meant a protein having at least about 35 %, usually at least about 40% and more usually at least about 60 % amino acid sequence identity with one of the specific human MPTS proteins as identified above (i.e. with a protein having the amino acid sequence of SEQ ID NOS:01, 03, 05 or 07), where sequence identity is determined as described *supra*.

The proteins of the subject invention are present in a non-naturally occurring environment, e.g. they are separated from their naturally occurring environment. In certain embodiments, the subject proteins are present in a composition that is enriched for the subject protein as compared to its naturally occurring environment. For example, purified protein is provided, where by purified is meant that the protein is present in a composition that is substantially free of non-MPTS proteins, where by substantially free is meant that less than 90 %, usually less than 60 % and more usually less than 50 % of the composition is made up of non- MPTS proteins. The proteins of the subject invention may also be present as an isolate, by which is meant that the protein is substantially free of other proteins and other naturally occurring biologic molecules, such as oligosaccharides, polynucleotides and fragments thereof, and the like, where the term "substantially free" in this instance means that less than 70 %,

usually less than 60% and more usually less than 50 % of the composition containing the isolated protein is some other naturally occurring biological molecule. In certain embodiments, the proteins are present in substantially pure form, where by "substantially pure form" is meant at least 95%, usually at least 97% and more usually at least 99% pure.

In addition to the naturally occurring proteins, polypeptides which vary from the naturally occurring proteins are also provided, e.g. MPTS polypeptides. By MPTS polypeptide is meant an amino acid sequence encoded by an open reading frame (ORF) of the gene encoding the MPTS, described in greater detail below, including the full length protein and fragments thereof, particularly biologically active fragments and/or fragments corresponding to functional domains, e.g. protease domain, thrombospondin domain, and the like; and including fusions of the subject polypeptides to other proteins or parts thereof. Fragments of interest will typically be at least about 10 aa in length, usually at least about 50 aa in length, and may be as long as 300 aa in length or longer, but will usually not exceed about 1000 aa in length, where the fragment will have a stretch of amino acids that is identical to the subject protein of at least about 10 aa, and usually at least about 15 aa, and in many embodiments at least about 50 aa in length. Where the fragment is an MPTS-15 fragment, it preferably includes at least a substantial portion of the protease domain of the wild type protein, where by substantial amount is at least 50%, usually at least 60 % and more usually at least 70 % of the sequence of this domain of the MPTS-15 protein. For example, the MPTS-15 fragment generally includes a sequence which , upon alignment with the sequence of residues from the protease domain of the wild type sequence, shows an identity with the aligned region of the wild type sequence of this domain of at least about 50%, usually at least about 60% and more usually at least about 70%, wherein in many embodiments the percent identity may be much higher, e.g. 75, 80, 85, 90 or 95% or higher, e.g. 99%.

carriage and the like. The subject proteins may also be derived from synthetic means.

e.g. by expressing a recombinant gene encoding protein of interest in a suitable host, as described in greater detail below. Any convenient protein purification procedures may be employed, where suitable protein purification methodologies are described in Guide to Protein Purification, (Deuthser ed.) (Academic Press, 1990). For example, a lysate may prepared from the original source, e.g. chondrocytes or the expression host, and purified using HPLC, exclusion chromatography, gel electrophoresis, affinity chromatography, and the like.

Also provided are nucleic acid compositions encoding MPTS proteins or fragments thereof, as well as the MPTS homologues of the present invention. By nucleic acid composition is meant a composition comprising a sequence of DNA having an open reading frame that encodes an MPTS polypeptide of the subject invention, i.e. an *mpts* gene, and is capable, under appropriate conditions, of being expressed as MPTS. Also encompassed in this term are nucleic acids that are homologous or substantially similar or identical to the nucleic acids encoding MPTS proteins. Thus, the subject invention provides genes encoding the human MPTS proteins of the subject invention and homologs thereof. The human MPTS15 gene is shown in Fig. 1A, where the sequence shown in Fig. 1A is identified as SEQ ID NO:02, *infra*. The human MPTS10 gene is shown in Fig. 2A, where the sequence shown in Fig. 2A is identified as SEQ ID NO:04, *infra*. The human MPTS19 gene is shown in Fig. 3A, where the sequence shown in Fig. 3A is identified as SEQ ID NO:06, *infra*. The human MPTS20 gene is shown in Fig. 4A, where the sequence shown in Fig. 4A is identified as SEQ ID NO:08, *infra*.

The source of homologous genes may be any species, e.g., primate species, particularly human, rodents, such as rats and mice, canines, felines, bovines, ovines, equines, yeast, nematodes, etc. Between mammalian species, *e.g.*, human and mouse, homologs have substantial sequence similarity, *e.g.* at least 75% sequence identity, usually at least 90%, more usually at least 95% between nucleotide sequences. Sequence similarity is calculated based on a reference sequence, which may be a subset of a larger sequence, such as a conserved motif, coding region, flanking region, *etc.* A reference sequence will usually be at least about 18 nt long, more usually at least about 30 nt long, and may extend to the complete sequence that is being compared. Algorithms for sequence analysis are known in the art, such as BLAST, described in Altschul *et al.*

(1990), *J. Mol. Biol.* 215:403-10 (using default settings, i.e. parameters *w*=4 and *T*=17). The sequences provided herein are essential for recognizing MPTS-, including aggrecanase-, related and homologous proteins, and the nucleic acids encoding the same, in database searches. Of particular interest in certain embodiments are nucleic acids of substantially the same length as the nucleic acids identified as SEQ ID NO:02, 04, 06 and 08 and have sequence identity to one of these sequences of at least about 90%, usually at least about 95% and more usually at least about 99% over the entire length of the nucleic acid.

Nucleic acids encoding the proteins and polypeptides of the subject invention may be cDNA or genomic DNA or a fragment thereof. The term "MPTS gene" shall be intended to mean the open reading frame encoding specific MPTS proteins and polypeptides, and introns, as well as adjacent 5′ and 3′ non-coding nucleotide sequences involved in the regulation of expression, up to about 20 kb beyond the coding region, but possibly further in either direction. The gene may be introduced into an appropriate vector for extrachromosomal maintenance or for integration into a host genome.

The term "cDNA" as used herein is intended to include all nucleic acids that share the arrangement of sequence elements found in native mature mRNA species, where sequence elements are exons and 5′ and 3′ non-coding regions. Normally mRNA species have contiguous exons, with the intervening introns, when present, being removed by nuclear RNA splicing, to create a continuous open reading frame encoding an MPTS protein.

A genomic sequence of interest comprises the nucleic acid present between the initiation codon and the stop codon, as defined in the listed sequences, including all of the introns that are normally present in a native chromosome. It may further include 5′ and 3′ untranslated sequences found in association with the coding region.

Where the coding region is interrupted by non-coding introns, the exons and intervening sequences within the transcribed region, the genomic DNA may be isolated as a

fragment of 100 kbp or smaller; and substantially free of flanking chromosomal sequence. The genomic DNA flanking the coding region, either 3' or 5', or internal regulatory sequences as sometimes found in introns, contains sequences required for proper tissue and stage specific expression.

The nucleic acid compositions of the subject invention may encode all or a part of the subject MPTS protein. Double or single stranded fragments may be obtained from the DNA sequence by chemically synthesizing oligonucleotides in accordance with conventional methods, by restriction enzyme digestion, by PCR amplification, *etc.* For the most part, DNA fragments will be of at least 15 nt, usually at least 18 nt or 25 nt, and may be at least about 50 nt.

The subject genes are isolated and obtained in substantial purity, generally as other than an intact chromosome. Usually, the DNA will be obtained substantially free of other nucleic acid sequences that do not include an MPTS gene sequence or fragment thereof, generally being at least about 50%, usually at least about 90% pure and are typically "recombinant", *i.e.* flanked by one or more nucleotides with which it is not normally associated on a naturally occurring chromosome.

In addition to the plurality of uses described in greater detail in following sections, the subject nucleic acid compositions find use in the preparation of all or a portion of the MPTS polypeptides, as described above. The provided polynucleotide (*e.g.*, a polynucleotide having a sequence of SEQ ID NO:02, 04, 06 or 08), the corresponding cDNA, or the full-length gene is used to express a partial or complete gene product. Constructs of polynucleotides having a sequences of SEQ ID NOs: 02, 04, 06 or 08 can be generated synthetically. Alternatively, single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides is described by, *e.g.*, Stemmer *et al., Gene (Amsterdam)* (1995) *164(1)*:49-53. In this method, assembly PCR (the synthesis of long DNA sequences from large numbers of oligodeoxy-ribonucleotides (oligos)) is described. The method is derived from DNA shuffling (Stemmer, *Nature* (1994) *370*:389-391), and does not rely on DNA ligase, but instead relies on DNA polymerase to build increasingly longer DNA fragments during the assembly process. Appropriate polynucleotide constructs are purified using standard

recombinant DNA techniques as described in, for example, Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual, 2nd Ed.*, (1989) Cold Spring Harbor Press, Cold Spring Harbor, NY, and under current regulations described in United States Dept. of HHS, National Institute of Health (NIH) Guidelines for Recombinant DNA Research.

Polynucleotide molecules comprising a polynucleotide sequence provided herein are propagated by placing the molecule in a vector. Viral and non-viral vectors are used, including plasmids. The choice of plasmid will depend on the type of cell in which propagation is desired and the purpose of propagation. Certain vectors are useful for amplifying and making large amounts of the desired DNA sequence. Other vectors are suitable for expression in cells in culture. Still other vectors are suitable for transfer and expression in cells in a whole animal or person. The choice of appropriate vector is well within the skill of the art. Many such vectors are available commercially. The partial or full-length polynucleotide is inserted into a vector typically by means of DNA ligase attachment to a cleaved restriction enzyme site in the vector. Alternatively, the desired nucleotide sequence can be inserted by homologous recombination in vivo. Typically this is accomplished by attaching regions of homology to the vector on the flanks of the desired nucleotide sequence. Regions of homology are added by ligation of oligonucleotides, or by polymerase chain reaction using primers comprising both the region of homology and a portion of the desired nucleotide sequence, for example.

For expression, an expression cassette or system may be employed. The gene product encoded by a polynucleotide of the invention is expressed in any convenient expression system, including, for example, bacterial, yeast, insect, amphibian and mammalian systems. Suitable vectors and host cells are described in U.S. Patent No. 5,654,173. In the expression vector, an MPTS encoding polynucleotide, e.g. as set forth in SEQ ID NO: 02, 04, 06 or 08, is linked to a regulatory sequence as appropriate to obtain the desired expression properties. These can include promoters (attached either at the 5' end of the sense strand or at the 3' end of the antisense strand), enhancers,

attached to the desired nucleotide sequence using the techniques described above for

linkage to vectors. Any techniques known in the art can be used. In other words, the expression vector will provide a transcriptional and translational initiation region, which may be inducible or constitutive, where the coding region is operably linked under the transcriptional control of the transcriptional initiation region, and a transcriptional and translational termination region. These control regions may be native to the subject MPTS gene, or may be derived from exogenous sources.

Expression vectors generally have convenient restriction sites located near the promoter sequence to provide for the insertion of nucleic acid sequences encoding heterologous proteins. A selectable marker operative in the expression host may be present. Expression vectors may be used for the production of fusion proteins, where the exogenous fusion peptide provides additional functionality, i.e. increased protein synthesis, stability, reactivity with defined antisera, an enzyme marker, e.g. β-galactosidase, etc.

Expression cassettes may be prepared comprising a transcription initiation region, the gene or fragment thereof, and a transcriptional termination region. Of particular interest is the use of sequences that allow for the expression of functional epitopes or domains, usually at least about 8 amino acids in length, more usually at least about 15 amino acids in length, to about 25 amino acids, and up to the complete open reading frame of the gene. After introduction of the DNA, the cells containing the construct may be selected by means of a selectable marker, the cells expanded and then used for expression.

The MPTS proteins and polypeptides may be expressed in prokaryotes or eukaryotes in accordance with conventional ways, depending upon the purpose for expression. For large scale production of the protein, a unicellular organism, such as *E. coli, B. subtilis, S. cerevisiae*, insect cells in combination with baculovirus vectors, or cells of a higher organism such as vertebrates, particularly mammals, e.g. COS 7 cells, HEK 293, CHO, Xenopus Oocytes, etc., may be used as the expression host cells. In some situations, it is desirable to express the gene in eukaryotic cells, where the expressed protein will benefit from native folding and post-translational modifications. Small peptides can also be synthesized in the laboratory. Polypeptides that are subsets of the

complete protein sequence may be used to identify and investigate parts of the protein important for function.

Specific expression systems of interest include bacterial, yeast, insect cell and mammalian cell derived expression systems. Representative systems from each of these categories is are provided below

Bacteria. Expression systems in bacteria include those described in Chang *et al.*, *Nature* (1978) *275*:615; Goeddel *et al.*, *Nature* (1979) *281*:544; Goeddel *et al.*, *Nucleic Acids Res.* (1980) *8*:4057; EP 0 036,776; U.S. Patent No. 4,551,433; DeBoer *et al.*, *Proc. Natl. Acad. Sci. (USA)* (1983) *80*:21-25; and Siebenlist *et al.*, *Cell* (1980) *20*:269.

Yeast. Expression systems in yeast include those described in Hinnen *et al.*, *Proc. Natl. Acad. Sci. (USA)* (1978) *75* 1929; Ito *et al.*, *J. Bacteriol.* (1983) *153*:163; Kurtz *et al.*, *Mol. Cell. Biol.* (1986) *6* 142; Kunze *et al.*, *J. Basic Microbiol.* (1985) *25*:141; Gleeson *et al.*, *J. Gen. Microbiol.* (1986) *132* 3459; Roggenkamp *et al.*, *Mol. Gen. Genet.* (1986) *202*:302; Das *et al.*, *J. Bacteriol.* (1984) *158*:1165; De Louvencourt *et al.*, *J. Bacteriol.* (1983) *154*:737; Van den Berg *et al.*, *Bio/Technology* (1990) *8*:135; Kunze *et al.*, *J. Basic Microbiol.* (1985) *25*:141; Cregg *et al.*, *Mol. Cell. Biol.* (1985) 5:3376; U.S. Patent Nos. 4,837,148 and 4,929,555; Beach and Nurse, *Nature* (1981) *300*:706; Davidow *et al.*, *Curr. Genet.* (1985) *10*:380; Gaillardin *et al.*, *Curr. Genet.* (1985) *10*:49; Ballance *et al.*, *Biochem. Biophys. Res. Commun.* (1983) *112*:284-289; Tilburn *et al.*, *Gene* (1983) *26*:205-221; Yelton *et al.*, *Proc. Natl. Acad. Sci. (USA)* (1984) *81*:1470-1474; Kelly and Hynes, *EMBO J.* (1985) 4:475479; EP 0 244,234; and WO 91/00357.

Insect Cells. Expression of heterologous genes in insects is accomplished as described in U.S. Patent No. 4,745,051; Friesen *et al.*, "The Regulation of Baculovirus Gene Expression", in: *The Molecular Biology Of Baculoviruses* (1986) (W. Doerfler, ed.); EP 0 127,839; EP 0 155,476; and Vlak *et al.*, *J. Gen. Virol.* (1988) 69:765-776; Miller *et al.*, *Ann. Rev. Microbiol.* (1988) *42*:177; Carbonell *et al.*, *Gene* (1988) *73*:409; Maeda *et al.*, *Nature* (1985) *315*:592-594; Lebacq-Verheyden *et al.*, *Mol. Cell. Biol.* (1988) 8:3129;

al., *Bio/Technology* (1988) *6*:47-55, Miller *et al.*, *Generic Engineering* (1986) *8*:277-279, and Maeda *et al.*, *Nature* (1985) *315*:592-594.

Mammalian Cells. Mammalian expression is accomplished as described in Dijkema *et al.*, *EMBO J.* (1985) *4*:761, Gorman *et al.*, *Proc. Natl. Acad. Sci. (USA)* (1982) *79*:6777, Boshart *et al.*, *Cell* (1985) *41*:521 and U.S. Patent No. 4,399,216. Other features of mammalian expression are facilitated as described in Ham and Wallace, *Meth. Enz.* (1979) *58*:44, Barnes and Sato, *Anal. Biochem.* (1980) *102*:255, U.S. Patent Nos. 4,767,704, 4,657,866, 4,927,762, 4,560,655, WO 90/103430, WO 87/00195, and U.S. RE 30,985.

When any of the above host cells, or other appropriate host cells or organisms, are used to replicate and/or express the polynucleotides or nucleic acids of the invention, the resulting replicated nucleic acid, RNA, expressed protein or polypeptide, is within the scope of the invention as a product of the host cell or organism. The product is recovered by any appropriate means known in the art.

Once the gene corresponding to a selected polynucleotide is identified, its expression can be regulated in the cell to which the gene is native. For example, an endogenous gene of a cell can be regulated by an exogenous regulatory sequence as disclosed in U.S. Patent No. 5,641,670.

The subject polypeptide and nucleic acid compositions find use in a variety of different applications, including general applications, diagnostic applications, and therapeutic agent screening/discovery/ preparation applications, as well as in therapeutic compositions and methods employing the same.

The subject nucleic acid compositions find use in a variety of general applications. General applications of interest include: the identification of MPTS homologs; as a source of novel promoter elements; the identification of MPTS expression regulatory factors; as probes and primers in hybridization applications, e.g. PCR; the identification of expression patterns in biological specimens; the preparation of cell or animal models for MPTS function; the preparation of *in vitro* models for MPTS function; etc.

Homologs of the subject genes are identified by any of a number of methods. A fragment of the provided cDNA may be used as a hybridization probe against a cDNA library from the target organism of interest, where low stringency conditions are used. The probe may be a large fragment, or one or more short degenerate primers. Nucleic acids having sequence similarity are detected by hybridization under low stringency conditions, for example, at 50°C and 6xSSC (0.9 M sodium chloride/0.09 M sodium citrate) and remain bound when subjected to washing at 55°C in 1xSSC (0.15 M sodium chloride/0.015 M sodium citrate). Sequence identity may be determined by hybridization under stringent conditions, for example, at 50°C or higher and 0.1xSSC (15 mM sodium chloride/01.5 mM sodium citrate). Nucleic acids having a region of substantial identity to the provided sequences, e.g. allelic variants, genetically altered versions of the gene, etc., bind to the provided sequences under stringent hybridization conditions. By using probes, particularly labeled probes of DNA sequences, one can isolate homologous or related genes.

The sequence of the 5' flanking region may be utilized for promoter elements, including enhancer binding sites, that provide for developmental regulation in tissues where the subject MPTS gene is expressed. The tissue specific expression is useful for determining the pattern of expression, and for providing promoters that mimic the native pattern of expression. Naturally occurring polymorphisms in the promoter region are useful for determining natural variations in expression, particularly those that may be associated with disease.

Alternatively, mutations may be introduced into the promoter region to determine the effect of altering expression in experimentally defined systems. Methods for the identification of specific DNA motifs involved in the binding of transcriptional factors are known in the art, e.g. sequence similarity to known binding motifs, gel retardation studies, *etc.* For examples, see Blackwell *et al.* (1995), *Mol. Med.* 1:194-205;

The regulatory sequences may be used to identify *cis* acting sequences required for transcriptional or translational regulation of MPTS gene expression, especially in different tissues or stages of development, and to identify *cis* acting sequences and *trans*-acting factors that regulate or mediate MPTS gene expression. Such transcription or translational control regions may be operably linked to an MPTS gene in order to promote expression of wild type or altered MPTS or other proteins of interest in cultured cells, or in embryonic, fetal or adult tissues, and for gene therapy.

Small DNA fragments are useful as primers for PCR, hybridization screening probes, *etc.* Larger DNA fragments, *i.e.* greater than 100 nt are useful for production of the encoded polypeptide, as described in the previous section. For use in geometric amplification reactions, such as geometric PCR, a pair of primers will be used. The exact composition of the primer sequences is not critical to the invention, but for most applications the primers will hybridize to the subject sequence under stringent conditions, as known in the art. It is preferable to choose a pair of primers that will generate an amplification product of at least about 50 nt, preferably at least about 100 nt. Algorithms for the selection of primer sequences are generally known, and are available in commercial software packages. Amplification primers hybridize to complementary strands of DNA, and will prime towards each other.

The DNA may also be used to identify expression of the gene in a biological specimen. The manner in which one probes cells for the presence of particular nucleotide sequences, as genomic DNA or RNA, is well established in the literature. Briefly, DNA or mRNA is isolated from a cell sample. The mRNA may be amplified by RT-PCR, using reverse transcriptase to form a complementary DNA strand, followed by polymerase chain reaction amplification using primers specific for the subject DNA sequences. Alternatively, the mRNA sample is separated by gel electrophoresis, transferred to a suitable support, *e.g.* nitrocellulose, nylon, *etc.*, and then probed with a fragment of the subject DNA as a probe. Other techniques, such as oligonucleotide ligation assays, *in situ* hybridizations, and hybridization to DNA probes arrayed on a solid chip may also find use. Detection of mRNA hybridizing to the subject sequence is indicative of MPTS gene expression in the sample.

The sequence of an MPTS gene, including flanking promoter regions and coding regions, may be mutated in various ways known in the art to generate targeted changes in promoter strength, sequence of the encoded protein, *etc.* The DNA sequence or protein product of such a mutation will usually be substantially similar to the sequences provided herein, *i.e.* will differ by at least one nucleotide or amino acid, respectively, and may differ by at least two but not more than about ten nucleotides or amino acids. The sequence changes may be substitutions, insertions, deletions, or a combination thereof. Deletions may further include larger changes, such as deletions of a domain or exon. Other modifications of interest include epitope tagging, *e.g.* with the FLAG system, HA, *etc.* For studies of subcellular localization, fusion proteins with green fluorescent proteins (GFP) may be used.

Techniques for *in vitro* mutagenesis of cloned genes are known. Examples of protocols for site specific mutagenesis may be found in Gustin *et al. (1993)*, *Biotechniques* 14:22; Barany (1985), *Gene* 37:111-23, Colicelli *et al. (1985)*, *Mol. Gen. Genet.* 199:537-9; and Prentki *et al.* (1984), *Gene* 29:303-13. Methods for site specific mutagenesis can be found in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, CSH Press 1989, pp. 15.3-15.108; Weiner *et al.* (1993), *Gene* 126:35-41; Sayers *et al.* (1992), *Biotechniques* 13:592-6; Jones and Winistorfer (1992), *Biotechniques* 12:528-30; Barton *et al.* (1990), *Nucleic Acids Res* 18:7349-55; Marotti and Tomich (1989), *Gene Anal. Tech.* 6:67-70; and Zhu (1989), *Anal Biochem* 177:120-4. Such mutated genes may be used to study structure-function relationships of an MPTS protein, or to alter properties of the protein that affect its function or regulation.

The subject nucleic acids can be used to generate transgenic, non-human animals or site specific gene modifications in cell lines. Transgenic animals may be made through homologous recombination, where the endogenous locus is altered. Alternatively, a nucleic acid construct is randomly integrated into the genome. Vectors for stable integration include plasmids, retroviruses and other animal viruses, YACs,

regulation. Of interest is the use of the subject genes to construct transgenic animal

models of MPTS related disease conditions, including aggrecanase related disease conditions, e.g. disease conditions associated with aggrecanase activity, such as arthritis. Thus, transgenic animal models of the subject invention include endogenous MPTS gene knockouts in which expression of endogenous MPTS is at least reduced if not eliminated, where such animals also typically express an MPTS peptide of the subject invention, e.g. the specific MPTS proteins of the subject invention or a fragment thereof. Where a nucleic acid having a sequence found in the human MPTS gene is introduced, the introduced nucleic acid may be either a complete or partial sequence of the MPTS gene. A detectable marker, such as *lac Z* may be introduced into the MPTS locus, where upregulation of gene expression will result in an easily detected change in phenotype. One may also provide for expression of the gene or variants thereof in cells or tissues where it is not normally expressed, at levels not normally present in such cells or tissues.

DNA constructs for homologous recombination will comprise at least a portion of the an MPTS gene of the subject invention, wherein the gene has the desired genetic modification(s), and includes regions of homology to the target locus. DNA constructs for random integration need not include regions of homology to mediate recombination. Conveniently, markers for positive and negative selection are included. Methods for generating cells having targeted gene modifications through homologous recombination are known in the art. For various techniques for transfecting mammalian cells, see Keown *et al.* (1990), *Meth. Enzymol.* 185:527-537.

For embryonic stem (ES) cells, an ES cell line may be employed, or embryonic cells may be obtained freshly from a host, e.g. mouse, rat, guinea pig, *etc.* Such cells are grown on an appropriate fibroblast-feeder layer or grown in the presence of leukemia inhibiting factor (LIF). When ES or embryonic cells have been transformed, they may be used to produce transgenic animals. After transformation, the cells are plated onto a feeder layer in an appropriate medium. Cells containing the construct may be detected by employing a selective medium. After sufficient time for colonies to grow, they are picked and analyzed for the occurrence of homologous recombination or integration of the construct. Those colonies that are positive may then be used for embryo manipulation and blastocyst injection. Blastocysts are obtained from 4 to 6 week old

superovulated females. The ES cells are trypsinized, and the modified cells are injected into the blastocoel of the blastocyst. After injection, the blastocysts are returned to each uterine horn of pseudopregnant females. Females are then allowed to go to term and the resulting offspring screened for the construct. By providing for a different phenotype of the blastocyst and the genetically modified cells, chimeric progeny can be readily detected.

The chimeric animals are screened for the presence of the modified gene and males and females having the modification are mated to produce homozygous progeny. If the gene alterations cause lethality at some point in development, tissues or organs can be maintained as allogeneic or congenic grafts or transplants, or in *in vitro* culture. The transgenic animals may be any non-human mammal, such as laboratory animals, domestic animals, etc. The transgenic animals may be used in functional studies, drug screening, *etc.*, *e.g.* to determine the effect of a candidate drug on aggrecanase activity.

Also provided are methods of diagnosing disease states based on observed levels of an MPTS protein or the expression level of the gene in a biological sample of interest. Samples, as used herein, include biological fluids such as blood, cerebrospinal fluid, tears, saliva, lymph, dialysis fluid, and the like; organ or tissue culture derived fluids; and fluids extracted from physiological tissues. Also included in the term are derivatives and fractions of such fluids. The cells may be dissociated, in the case of solid tissues, or tissue sections may be analyzed. Alternatively a lysate of the cells may be prepared.

A number of methods are available for determining the expression level of a gene or protein in a particular sample. Diagnosis may be performed by a number of methods to determine the absence or presence or altered amounts of normal or abnormal MPTS in a patient sample. For example, detection may utilize staining of cells or histological sections with labeled antibodies, performed in accordance with conventional methods. Cells are permeabilized to stain cytoplasmic molecules. The

other labels for direct detection. Alternatively, a second stage antibody or reagent is

used to amplify the signal. Such reagents are well known in the art. For example, the primary antibody may be conjugated to biotin, with horseradish peroxidase-conjugated avidin added as a second stage reagent. Alternatively, the secondary antibody conjugated to a fluorescent compound, *e.g.* fluorescein, rhodamine, Texas red, *etc.* Final detection uses a substrate that undergoes a color change in the presence of the peroxidase. The absence or presence of antibody binding may be determined by various methods, including flow cytometry of dissociated cells, microscopy, radiography, scintillation counting, *etc.*

Alternatively, one may focus on the expression of the MPTS gene. Biochemical studies may be performed to determine whether a sequence polymorphism in an MPTS coding region or control regions is associated with disease. Disease associated polymorphisms may include deletion or truncation of the gene, mutations that alter expression level, that affect the activity of the protein, *etc.*

Changes in the promoter or enhancer sequence that may affect expression levels of MPTS can be compared to expression levels of the normal allele by various methods known in the art. Methods for determining promoter or enhancer strength include quantitation of the expressed natural protein; insertion of the variant control element into a vector with a reporter gene such as β-galactosidase, luciferase, chloramphenicol acetyltransferase, *etc.* that provides for convenient quantitation; and the like.

A number of methods are available for analyzing nucleic acids for the presence of a specific sequence, *e.g.* a disease associated polymorphism. Where large amounts of DNA are available, genomic DNA is used directly. Alternatively, the region of interest is cloned into a suitable vector and grown in sufficient quantity for analysis. Cells that express an MPTS protein may be used as a source of mRNA, which may be assayed directly or reverse transcribed into cDNA for analysis. The nucleic acid may be amplified by conventional techniques, such as the polymerase chain reaction (PCR), to provide sufficient amounts for analysis. The use of the polymerase chain reaction is described in Saiki, *et al.* (1985), *Science* 239:487, and a review of techniques may be found in Sambrook, *et al.* Molecular Cloning: A Laboratory Manual, CSH Press 1989, pp.14.2–14.33. Alternatively, various methods are known in the art that utilize

oligonucleotide ligation as a means of detecting polymorphisms, for examples see Riley *et al.* (1990), *Nucl. Acids Res.* 18:2887-2890; and Delahunty *et al.* (1996), *Am. J. Hum. Genet.* 58:1239-1246.

A detectable label may be included in an amplification reaction. Suitable labels include fluorochromes, *e.g.* fluorescein isothiocyanate (FITC), rhodamine, Texas Red, phycoerythrin, allophycocyanin, 6-carboxyfluorescein (6-FAM), 2′,7′-dimethoxy-4′,5′-dichloro-6-carboxyfluorescein (JOE), 6-carboxy-X-rhodamine (ROX), 6-carboxy-2′,4′,7′,4,7-hexachlorofluorescein (HEX), 5-carboxyfluorescein (5-FAM) or N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA), radioactive labels, *e.g.* ^{32}P, ^{35}S, ^{3}H; *etc.* The label may be a two stage system, where the amplified DNA is conjugated to biotin, haptens, *etc.* having a high affinity binding partner, *e.g.* avidin, specific antibodies, *etc.*, where the binding partner is conjugated to a detectable label. The label may be conjugated to one or both of the primers. Alternatively, the pool of nucleotides used in the amplification is labeled, so as to incorporate the label into the amplification product.

The sample nucleic acid, *e.g.* amplified or cloned fragment, is analyzed by one of a number of methods known in the art. The nucleic acid may be sequenced by dideoxy or other methods, and the sequence of bases compared to a wild-type gene sequence. Hybridization with the variant sequence may also be used to determine its presence, by Southern blots, dot blots, *etc.* The hybridization pattern of a control and variant sequence to an array of oligonucleotide probes immobilized on a solid support, as described in US 5,445,934, or in WO 95/35505, may also be used as a means of detecting the presence of variant sequences. Single strand conformational polymorphism (SSCP) analysis, denaturing gradient gel electrophoresis (DGGE), and heteroduplex analysis in gel matrices are used to detect conformational changes created by DNA sequence variation as alterations in electrophoretic mobility. Alternatively, where a polymorphism creates or destroys a recognition site for a restriction endonuclease, the

Screening for mutations in MPTS may be based on the functional or antigenic characteristics of the protein. Protein truncation assays are useful in detecting deletions that may affect the biological activity of the protein. Various immunoassays designed to detect polymorphisms in MPTS proteins may be used in screening. Where many diverse genetic mutations lead to a particular disease phenotype, functional protein assays have proven to be effective screening tools. The activity of the encoded MPTS protein may be determined by comparison with the wild-type protein.

Diagnostic methods of the subject invention in which the level of MPTS gene expression is of interest will typically involve comparison of the MPTS nucleic acid abundance of a sample of interest with that of a control value to determine any relative differences, where the difference may be measured qualitatively and/or quantitatively, which differences are then related to the presence or absence of an abnormal MPTS gene expression pattern. A variety of different methods for determine the nucleic acid abundance in a sample are known to those of skill in the art, where particular methods of interest include those described in: Pietu et al., Genome Res. (June 1996) 6: 492-503; Zhao et al., Gene (April 24, 1995) 156: 207-213; Soares , Curr. Opin. Biotechnol. (October 1997) 8: 542-546; Raval, J. Pharmacol Toxicol Methods (November 1994) 32: 125-127; Chalifour et al., Anal. Biochem (February 1, 1994) 216: 299-304; Stolz & Tuan, Mol. Biotechnol. (December 19960 6: 225-230; Hong et al., Bioscience Reports (1982) 2: 907; and McGraw, Anal. Biochem. (1984) 143: 298. Also of interest are the methods disclosed in WO 97/27317, the disclosure of which is herein incorporated by reference.

The subject polypeptides find use in various screening assays designed to identify therapeutic agents. In vitro screening assays can be employed in which the activity of an MPTS polypeptide, e.g. the aggrecanase activity of an MPTS polypeptide, is assessed in the presence of a candidate therapeutic agent and compared to a control, i.e. the activity in the absence of the candidate therapeutic agent. Activity can be determined in a number of different ways, where activity may generally be determined as ability to cleave aggrecan or at least a fragment therefore, as well as a recombinant polypeptide, that includes the aggrecanase cleavage site, as described above. Such assays are described in U.S. Patent No. 5,872,209 and WO 99/05921 as well as Arner et al., J. Biol. Chem. (March 1999) 274: 6594-6601.

Also of interest in screening assays are non-human transgenic animals that express functional MPTS, where such animals are described above. In many embodiments, the animals lack the corresponding endogenous MPTS. In using such animals for screening applications, a test compound(s) is administered to the animal, and the resultant changes in phenotype, e.g. presence of aggrecan produced by cleavage of the Glu^{373}-Ala^{374} bond, are compared with a control.

Alternatively, in vitro models of MPTS binding activity may be measured in which binding events between MPTS and candidate MPTS modulatory agents are monitored.

A variety of other reagents may be included in the screening assays, depending on the particular screening protocols employed. These include reagents like salts, neutral proteins, e.g. albumin, detergents, etc that are used to facilitate optimal protein-protein binding and/or reduce non-specific or background interactions. Reagents that improve the efficiency of the assay, such as protease inhibitors, nuclease inhibitors, anti-microbial agents, *etc.* may be used.

A variety of different candidate therapeutic agents that serve as either MPTS agonists or antagonists may be screened by the above methods. Candidate agents encompass numerous chemical classes, though typically they are organic molecules, preferably small organic compounds having a molecular weight of more than 50 and less than about 2,500 daltons. Candidate agents comprise functional groups necessary for structural interaction with proteins, particularly hydrogen bonding, and typically include at least an amine, carbonyl, hydroxyl or carboxyl group, preferably at least two of the functional chemical groups. The candidate agents often comprise cyclical carbon or heterocyclic structures and/or aromatic or polyaromatic structures substituted with one or more of the above functional groups. Candidate agents are also found among biomolecules including peptides, saccharides, fatty acids, steroids, purines, pyrimidines,

synthetic or natural compounds. For example, numerous means are available for

random and directed synthesis of a wide variety of organic compounds and biomolecules, including expression of randomized oligonucleotides and oligopeptides. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts are available or readily produced. Additionally, natural or synthetically produced libraries and compounds are readily modified through conventional chemical, physical and biochemical means, and may be used to produce combinatorial libraries. Known pharmacological agents may be subjected to directed or random chemical modifications, such as acylation, alkylation, esterification, amidification, *etc.* to produce structural analogs.

Of particular interest in many embodiments are screening methods that identify agents that selectively modulate, e.g. inhibit, the subject MPTS enzyme and not other proteases.

The nucleic acid compositions of the subject invention also find use as therapeutic agents in situations where one wishes to enhance an MPTS activity in a host. The MPTS genes, gene fragments, or the encoded proteins or protein fragments are useful in gene therapy to treat disorders associated with MPTS defects, including aggrecanase defects. Expression vectors may be used to introduce the gene into a cell. Such vectors generally have convenient restriction sites located near the promoter sequence to provide for the insertion of nucleic acid sequences. Transcription cassettes may be prepared comprising a transcription initiation region, the target gene or fragment thereof, and a transcriptional termination region. The transcription cassettes may be introduced into a variety of vectors, *e.g.* plasmid; retrovirus, *e.g.* lentivirus; adenovirus; and the like, where the vectors are able to transiently or stably be maintained in the cells, usually for a period of at least about one day, more usually for a period of at least about several days to several weeks.

The gene or protein may be introduced into tissues or host cells by any number of routes, including viral infection, microinjection, or fusion of vesicles. Jet injection may also be used for intramuscular administration, as described by Furth *et al.* (1992), *Anal Biochem* 205:365-368. The DNA may be coated onto gold microparticles, and delivered intradermally by a particle bombardment device, or "gene gun" as described in

the literature (see, for example, Tang *et al.* (1992), *Nature* 356:152-154), where gold microprojectiles are coated with the DNA, then bombarded into skin cells.

The subject invention provides methods of modulating MPTS, and in many embodiments aggrecanase, activity in a cell, including methods of increasing MPTS activity (e.g. methods of enhancing), as well as methods of reducing or inhibiting MPTS activity, e.g. methods of stopping or limiting aggrecan cleavage. In such methods, an effective amount of a modulatory agent is contacted with the cell.

Also provided are methods of modulating, including enhancing and inhibiting, MPTS activity in a host. In such methods, an effective amount of active agent that modulates the activity of an MPTS protein *in vivo*, e.g. where the agent usually enhances or inhibits the target MPTS activity, is administered to the host. The active agent may be a variety of different compounds, including a naturally occurring or synthetic small molecule compound, an antibody, fragment or derivative thereof, an antisense composition, and the like.

Of particular interest in certain embodiments are agents that reduce MPTS activity, including agents that reduce aggrecanase activity, e.g. aggrecan cleavage, by at least about 10 fold, usually at least about 20 fold and more usually at least about 25 fold, as measure by the Assay described in Arner et al. (1999), *supra*. In many embodiments, of particular interest is the use of compounds that reduce aggrecanase activity by at least 100 fold, as compared to a control.

Also of interest is the use of agents that, while providing for reduced MPTS, including aggrecanase, activity, do not substantially reduce the activity of other proteinases, if at all. Thus, the agents in this embodiment are selective inhibitors of MPTS. An agent is considered to be selective if it provides for the above reduced aggrecanase activity, but substantially no reduced activity of at least one other

Naturally occurring or synthetic small molecule compounds of interest include numerous chemical classes, though typically they are organic molecules, preferably small organic compounds having a molecular weight of more than 50 and less than about 2,500 daltons. Candidate agents comprise functional groups necessary for structural interaction with proteins, particularly hydrogen bonding, and typically include at least an amine, carbonyl, hydroxyl or carboxyl group, preferably at least two of the functional chemical groups. The candidate agents often comprise cyclical carbon or heterocyclic structures and/or aromatic or polyaromatic structures substituted with one or more of the above functional groups. Candidate agents are also found among biomolecules including peptides, saccharides, fatty acids, steroids, purines, pyrimidines, derivatives, structural analogs or combinations thereof.

Also of interest as active agents are antibodies that at least reduce, if not inhibit, the target MPTS, e.g. aggrecanase, activity in the host. Suitable antibodies are obtained by immunizing a host animal with peptides comprising all or a portion of the target protein, e.g. MPTS-15, MPTS-19 or MPTS-20. Suitable host animals include mouse, rat sheep, goat, hamster, rabbit, *etc.* The origin of the protein immunogen may be mouse, human, rat, monkey *etc.* The host animal will generally be a different species than the immunogen, *e.g.* human MPTS used to immunize mice, *etc.*

The immunogen may comprise the complete protein, or fragments and derivatives thereof. Preferred immunogens comprise all or a part of MPTS, where these residues contain the post-translation modifications, such as glycosylation, found on the native target protein. Immunogens comprising the extracellular domain are produced in a variety of ways known in the art, *e.g.* expression of cloned genes using conventional recombinant methods, isolation from HEC, *etc.*

For preparation of polyclonal antibodies, the first step is immunization of the host animal with the target protein, where the target protein will preferably be in substantially pure form, comprising less than about 1% contaminant. The immunogen may comprise the complete target protein, fragments or derivatives thereof. To increase the immune response of the host animal, the target protein may be combined with an adjuvant, where suitable adjuvants include alum, dextran, sulfate, large polymeric

anions, oil & water emulsions, e.g. Freund's adjuvant, Freund's complete adjuvant, and the like. The target protein may also be conjugated to synthetic carrier proteins or synthetic antigens. A variety of hosts may be immunized to produce the polyclonal antibodies. Such hosts include rabbits, guinea pigs, rodents, e.g. mice, rats, sheep, goats, and the like. The target protein is administered to the host, usually intradermally, with an initial dosage followed by one or more, usually at least two, additional booster dosages. Following immunization, the blood from the host will be collected, followed by separation of the serum from the blood cells. The Ig present in the resultant antiserum may be further fractionated using known methods, such as ammonium salt fractionation, DEAE chromatography, and the like.

Monoclonal antibodies are produced by conventional techniques. Generally, the spleen and/or lymph nodes of an immunized host animal provide a source of plasma cells. The plasma cells are immortalized by fusion with myeloma cells to produce hybridoma cells. Culture supernatant from individual hybridomas is screened using standard techniques to identify those producing antibodies with the desired specificity. Suitable animals for production of monoclonal antibodies to the human protein include mouse, rat, hamster, *etc.* To raise antibodies against the mouse protein, the animal will generally be a hamster, guinea pig, rabbit, etc. The antibody may be purified from the hybridoma cell supernatants or ascites fluid by conventional techniques, *e.g.* affinity chromatography using MPTS bound to an insolubie support, protein A sepharose, *etc.* Therefore it is an object of the present invention to provide monoclonal antibodies binding specifically to the MPTS proteins of the present invention, more specifically such antibodies which inhibit aggrecanase activity and such antibodies which are human or humanized ones.

The antibody may be produced as a single chain, instead of the normal multimeric structure. Single chain antibodies are described in Jost *et al.* (1994) J.B.C. 269:26267-73, and others. DNA sequences encoding the variable region of the heavy

retains the specificity and affinity of the original antibody

For *in vivo* use, particularly for injection into humans, it is desirable to decrease the antigenicity of the antibody. An immune response of a recipient against the blocking agent will potentially decrease the period of time that the therapy is effective. Methods of humanizing antibodies are known in the art. The humanized antibody may be the product of an animal having transgenic human immunoglobulin constant region genes (see for example WO 90/10077 and WO 90/04036). Alternatively, the antibody of interest may be engineered by recombinant DNA techniques to substitute the CH1, CH2, CH3, hinge domains, and/or the framework domain with the corresponding human sequence (see WO 92/02190).

The use of Ig cDNA for construction of chimeric immunoglobulin genes is known in the art (Liu *et al.* (1987) P.N.A.S. 84:3439 and (1987) J. Immunol. 139:3521). mRNA is isolated from a hybridoma or other cell producing the antibody and used to produce cDNA. The cDNA of interest may be amplified by the polymerase chain reaction using specific primers (U.S. Patent Nos. 4,683,195 and 4,683,202). Alternatively, a library is made and screened to isolate the sequence of interest. The DNA sequence encoding the variable region of the antibody is then fused to human constant region sequences. The sequences of human constant regions genes may be found in Kabat *et al.* (1991) Sequences of Proteins of Immunological Interest, N.I.H. publication no. 91-3242. Human C region genes are readily available from known clones. The choice of isotype will be guided by the desired effector functions, such as complement fixation, or activity in antibody-dependent cellular cytotoxicity. Preferred isotypes are IgG1, IgG3 and IgG4. Either of the human light chain constant regions, kappa or lambda, may be used. The chimeric, humanized antibody is then expressed by conventional methods.

In yet other embodiments, the antibodies may be fully human antibodies. For example, xenogeneic antibodies which are identical to human antibodies may be employed. By xenogenic human antibodies is meant antibodies that are the same has human antibodies, i.e. they are fully human antibodies, with exception that they are produced using a non-human host which has been genetically engineered to express human antibodies, e.g. WO 98/50433; WO 98,24893 and WO 99/53049.

Antibody fragments, such as Fv, $F(ab')_2$ and Fab may be prepared by cleavage of the intact protein, *e.g.* by protease or chemical cleavage. Alternatively, a truncated gene is designed. For example, a chimeric gene encoding a portion of the $F(ab')_2$ fragment would include DNA sequences encoding the CH1 domain and hinge region of the H chain, followed by a translational stop codon to yield the truncated molecule.

Consensus sequences of H and L J regions may be used to design oligonucleotides for use as primers to introduce useful restriction sites into the J region for subsequent linkage of V region segments to human C region segments. C region cDNA can be modified by site directed mutagenesis to place a restriction site at the analogous position in the human sequence.

Expression vectors include plasmids, retroviruses, YACs, EBV derived episomes, and the like. A convenient vector is one that encodes a functionally complete human CH or CL immunoglobulin sequence, with appropriate restriction sites engineered so that any VH or VL sequence can be easily inserted and expressed. In such vectors, splicing usually occurs between the splice donor site in the inserted J region and the splice acceptor site preceding the human C region, and also at the splice regions that occur within the human CH exons. Polyadenylation and transcription termination occur at native chromosomal sites downstream of the coding regions. The resulting chimeric antibody may be joined to any strong promoter, including retroviral LTRs, *e.g.* SV-40 early promoter, (Okayama *et al.* (1983) Mol. Cell. Bio. 3:280), Rous sarcoma virus LTR (Gorman *et al.* (1982) P.N.A.S. 79:6777), and moloney murine leukemia virus LTR (Grosschedl *et al.* (1985) Cell 41:885); native Ig promoters, *etc.*

In yet other embodiments of the invention, the active agent is an agent that modulates, and generally decreases or down regulates, the expression of the gene encoding the target protein in the host. For example, antisense molecules can be used to down regulate expression of MPTS in cells. The anti-sense reagent may be antisense [illegible] targeted gene, and inhibits expression of the targeted gene products. Antisense

molecules inhibit gene expression through various mechanisms, *e.g.* by reducing the amount of mRNA available for translation, through activation of RNAse H, or steric hindrance. One or a combination of antisense molecules may be administered, where a combination may comprise multiple different sequences

Antisense molecules may be produced by expression of all or a part of the target gene sequence in an appropriate vector, where the transcriptional initiation is oriented such that an antisense strand is produced as an RNA molecule. Alternatively, the antisense molecule is a synthetic oligonucleotide. Antisense oligonucleotides will generally be at least about 7, usually at least about 12, more usually at least about 20 nucleotides in length, and not more than about 500, usually not more than about 50, more usually not more than about 35 nucleotides in length, where the length is governed by efficiency of inhibition, specificity, including absence of cross-reactivity, and the like. It has been found that short oligonucleotides, of from 7 to 8 bases in length, can be strong and selective inhibitors of gene expression (see Wagner *et al.* (1996), *Nature Biotechnol.* 14:840-844).

A specific region or regions of the endogenous sense strand mRNA sequence is chosen to be complemented by the antisense sequence. Selection of a specific sequence for the oligonucleotide may use an empirical method, where several candidate sequences are assayed for inhibition of expression of the target gene in an *in vitro* or animal model. A combination of sequences may also be used, where several regions of the mRNA sequence are selected for antisense complementation.

Antisense oligonucleotides may be chemically synthesized by methods known in the art (see Wagner *et al.* (1993), *supra*, and Milligan *et al.*, *supra*.) Preferred oligonucleotides are chemically modified from the native phosphodiester structure, in order to increase their intracellular stability and binding affinity. A number of such modifications have been described in the literature, which alter the chemistry of the backbone, sugars or heterocyclic bases.

Among useful changes in the backbone chemistry are phosphorothioates; phosphorodithioates, where both of the non-bridging oxygens are substituted with

sulfur; phosphoroamidites; alkyl phosphotriesters and boranophosphates. Achiral phosphate derivatives include 3'-O'-5'-S-phosphorothioate, 3'-S-5'-O-phosphorothioate, 3'-CH2-5'-O-phosphonate and 3'-NH-5'-O-phosphoroamidate. Peptide nucleic acids replace the entire ribose phosphodiester backbone with a peptide linkage. Sugar modifications are also used to enhance stability and affinity. The α-anomer of deoxyribose may be used, where the base is inverted with respect to the natural β-anomer. The 2'-OH of the ribose sugar may be altered to form 2'-O-methyl or 2'-O-allyl sugars, which provides resistance to degradation without comprising affinity. Modification of the heterocyclic bases must maintain proper base pairing. Some useful substitutions include deoxyuridine for deoxythymidine; 5-methyl-2'-deoxycytidine and 5-bromo-2'-deoxycytidine for deoxycytidine. 5- propynyl-2'-deoxyuridine and 5-propynyl-2'-deoxycytidine have been shown to increase affinity and biological activity when substituted for deoxythymidine and deoxycytidine, respectively.

As an alternative to anti-sense inhibitors, catalytic nucleic acid compounds, *e.g.* ribozymes, anti-sense conjugates, *etc.* may be used to inhibit gene expression. Ribozymes may be synthesized *in vitro* and administered to the patient, or may be encoded on an expression vector, from which the ribozyme is synthesized in the targeted cell (for example, see International patent application WO 9523225, and Beigelman *et al.* (1995), *Nucl. Acids Res.* 23:4434-42). Examples of oligonucleotides with catalytic activity are described in WO 9506764. Conjugates of anti-sense ODN with a metal complex, *e.g.* terpyridylCu(II), capable of mediating mRNA hydrolysis are described in Bashkin *et al.* (1995), *Appl. Biochem. Biotechnol.* 54:43-56.

It is herefore an object of the present invention to provide a method of modulating MPTS activity in a host, said method comprising: administering an effective amount of an MPTS modulatory agent to said host, or more specifically such a method wherein said modulatory agent is a small molecule, an antibody, or a nucleic acid.

As mentioned above, an effective amount of the active agent is administered to

e.g. aggrecanase, activity, as measured by aggrecan cleavage product production, as compared to a control.

In the subject methods, the active agent(s) may be administered to the host using any convenient means capable of resulting in the desired modulation of MPTS activity, e.g. desired reduction in aggrecan cleavage product production. Thus, the agent can be incorporated into a variety of formulations for therapeutic administration. More particularly, the agents of the present invention can be formulated into pharmaceutical compositions by combination with appropriate, pharmaceutically acceptable carriers or diluents, and may be formulated into preparations in solid, semi-solid, liquid or gaseous forms, such as tablets, capsules, powders, granules, ointments, solutions, suppositories, injections, inhalants and aerosols.

As such, administration of the agents can be achieved in various ways, including oral, buccal, rectal, parenteral, intraperitoneal, intradermal, transdermal, intracheal,etc., administration.

In pharmaceutical dosage forms, the agents may be administered in the form of their pharmaceutically acceptable salts, or they may also be used alone or in appropriate association, as well as in combination, with other pharmaceutically active compounds.

For oral preparations, the agents can be used alone or in combination with appropriate additives to make tablets, powders, granules or capsules, for example, with conventional additives, such as lactose, mannitol, corn starch or potato starch; with binders, such as crystalline cellulose, cellulose derivatives, acacia, corn starch or gelatins; with disintegrators, such as corn starch, potato starch or sodium carboxymethylcellulose; with lubricants, such as talc or magnesium stearate; and if desired, with diluents, buffering agents, moistening agents, preservatives and flavoring agents.

The agents can be formulated into preparations for injection by dissolving, suspending or emulsifying them in an aqueous or nonaqueous solvent, such as vegetable or other similar oils, synthetic aliphatic acid glycerides, esters of higher aliphatic acids or

propylene glycol; and if desired, with conventional additives such as solubilizers, isotonic agents, suspending agents, emulsifying agents, stabilizers and preservatives.

The agents can be utilized in aerosol formulation to be administered via inhalation. The compounds of the present invention can be formulated into pressurized acceptable propellants such as dichlorodifluoromethane, propane, nitrogen and the like.

Furthermore, the agents can be made into suppositories by mixing with a variety of bases such as emulsifying bases or water-soluble bases. The compounds of the present invention can be administered rectally via a suppository. The suppository can include vehicles such as cocoa butter, carbowaxes and polyethylene glycols, which melt at body temperature, yet are solidified at room temperature.

Unit dosage forms for oral or rectal administration such as syrups, elixirs, and suspensions may be provided wherein each dosage unit, for example, teaspoonful, tablespoonful, tablet or suppository, contains a predetermined amount of the composition containing one or more inhibitors. Similarly, unit dosage forms for injection or intravenous administration may comprise the inhibitor(s) in a composition as a solution in sterile water, normal saline or another pharmaceutically acceptable carrier.

The term "unit dosage form," as used herein, refers to physically discrete units suitable as unitary dosages for human and animal subjects, each unit containing a predetermined quantity of compounds of the present invention calculated in an amount sufficient to produce the desired effect in association with a pharmaceutically acceptable diluent, carrier or vehicle. The specifications for the novel unit dosage forms of the present invention depend on the particular compound employed and the effect to be achieved, and the pharmacodynamics associated with each compound in the host.

agents, stabilizers, wetting agents and the like, are readily available to the public.

Where the agent is a polypeptide, polynucleotide, analog or mimetic thereof, e.g. antisense composition, it may be introduced into tissues or host cells by any number of routes, including viral infection, microinjection, or fusion of vesicles. Jet injection may also be used for intramuscular administration, as described by Furth *et al.* (1992), *Anal Biochem* 205:365-368. The DNA may be coated onto gold microparticles, and delivered intradermally by a particle bombardment device, or "gene gun" as described in the literature (see, for example, Tang *et al.* (1992), *Nature* 356:152-154), where gold microprojectiles are coated with the therapeutic DNA, then bombarded into skin cells.

Those of skill in the art will readily appreciate that dose levels can vary as a function of the specific compound, the severity of the symptoms and the susceptibility of the subject to side effects. Preferred dosages for a given compound are readily determinable by those of skill in the art by a variety of means.

The subject methods find use in the treatment of a variety of different disease conditions involving MPTS activity, including disease conditions involving aggrecanase activity. Of particular interest is the use of the subject methods to treat disease conditions characterized by the presence of aggrecan cleavage products, particularly 60 kDa aggrecan cleavage products having an ARGS N-terminus. Specific diseases that are characterized by the presence of such methods include: rheumatoid arthritis, osteo-arthritis, infectious arthritis, gouty arthritis, psoriatic arthritis, spondolysis, sports injury, joint trauma, pulmonary disease, fibrosis, and the like.

By treatment is meant at least an amelioration of the symptoms associated with the pathological condition afflicting the host, where amelioration is used in a broad sense to refer to at least a reduction in the magnitude of a parameter, *e.g.* symptom, associated with the pathological condition being treated, such as hyperphosphatemia. As such, treatment also includes situations where the pathological condition, or at least symptoms associated therewith, are completely inhibited, *e.g.* prevented from happening, or stopped, *e.g.* terminated, such that the host no longer suffers from the pathological condition, or at least the symptoms that characterize the pathological condition.

A variety of hosts are treatable according to the subject methods. Generally such hosts are "mammals" or "mammalian," where these terms are used broadly to describe organisms which are within the class mammalia, including the orders carnivore (*e.g.*, dogs and cats), rodentia (*e.g.*, mice, guinea pigs, and rats), and primates (*e.g.*, humans, chimpanzees, and monkeys). In many embodiments, the hosts will be humans.

Kits with unit doses of the active agent, usually in oral or injectable doses, are provided. In such kits, in addition to the containers containing the unit doses will be an informational package insert describing the use and attendant benefits of the drugs in treating pathological condition of interest. Preferred compounds and unit doses are those described herein above.

Finaly it is an object of the present invention:

(i) A method of screening to identify MPTS modulatory agents, said method comprising:
contacting an MPTS proteins as defined herein with a substrate in the presence of an potential modulatory agents; and
determing the effect of said modulatory agent on the activity of said protein

(ii) The method as defined in (i), wherein said substrate comprises a glu-ala bond.

(iii) The method as defined in (ii), wherein said substrate is aggrecan or a fragment thereof.

Furthermore, a method of treating a host suffering from a disease condition associated with MPTS activity specifically wherein said disease condition is characterized by the presence of aggrecan cleavage prducts, said method comprising:

[illegible] the treatment of a disease condition associated with MPTS activity.

specifically wherein said disease condition is characterized by the presence of aggrecan cleavage products, like arthritis.

Examples

Example 1

A nucleic acid array carrying 699 known metalloproteinase genes and novel ESTs available in public and proprietary databases was designed. These sequences on the array were selected by a search with a seed set of known metalloprotease protein sequences from all species. These protein sequences were used to find matching sequences in human nucleotide at the protein (codon) level. Redundant sequences were eliminated, remaining sequences assembled and clustered, and the unique set of 699 sequences were arrayed.

The resultant array was used to screen genes expressed in primary cultures of chondrocytes. A fair number of metalloproteinases known to be expressed by these cells were identified. However, a number of ESTs for novel proteins were also identified. Using these ESTs in subsequent database mining and PCR protocols, four different human MPTS proteins were identified, i.e. MPTS15, MPTS10, MPTS19 and MPTS20.

Example 2

Expression of MPTS-10

An example of a system for expression of mpts-10 is the COS-7 mammalian cell system. The nucleotide sequence that encodes mpts-10, including the secretion signal sequence, was ligated into a pcDNA3.1 plasmid (In Vitrogen, Carlsbad, CA, USA). Two micrograms of the resulting plasmid was combined with lipofectamine (Life Technologies, Rockville, MD, USA). The mixture was then added to COS-7 cells, which were grown in 6 well plates to a density of approximately 90% confluency. After 6 hours, fresh medium was added to the cells and after 24 hours the cells were washed and fresh serum free medium containing bovine aggrecan (0.1mg/ml, Sigma, St. Louis, MO, USA) was added. The cells were incubated for an additional 48 hours. Five hundred

microliters of culture fluid from each well was collected and concentrated ten fold. Two microliters of chondroitinase ABC and keratinase (10 u/ml, Sigma, St. Louis, MO, USA) was then added and the samples incubated overnight at 37C. The samples were then boiled in SDS-PAGE sample loading buffer, electrophoresed on a polyacryamide gel and transferred to a PVDF membrane. A Western blot using an antiserum against a neoepitope generated when aggrecanase cleaves aggrecan was then performed.

Another example of a system for expression of mpts-10 was the baculovirus expression system. The DNA sequence that contained the coding sequence for mpts-10 (including the sequences that code for the secretion signal sequence) and that had been cloned in the pcDNA3.1 vector was modified by PCR so that the coding sequence and the translational stop codon were flanked by the Not 1 (N-terminal side) and Sfi-1 (C-terminal side). The primer used for the N-terminal end was GATCGCGGCCGCTATGGTGGACACGTGGCCTCTATGGCTCC and the primer for the C-terminal end was TGAGGCCTTCAGGGCCGATCACTGTGCAGAGCACTCACCCCAT. After amplification using standard PCR methods, the fragment was digested with Not 1 and Sfi-1. The digested fragment was ligated into a vector pVL1392-U, which had also been digested with Not1 and Sfi-1. PVL1392-U is a derivation of the baculovirus transfer plasmid, pVL1392 (PharMingen, San Diego, CA USA) in which the multiple cloning site has been modified to contain Not-1 and Sfi-1. The overhangs generated by digestion with Not-1 and Sfi-1 were complementary to the overhangs generated in the Not 1 and Sfi 1 digested PCR amplified DNA. The ligated DNA was transformed into bacterial cells and a clone was selected that contained the plasmid and the correct mpts-10 sequence. This plasmid was produced and purified. The mpts-10 sequence was transferred into a baculovirus vector using standard techniques (*Baculovirus Expression Vectors: A Laboratory Manual* by David O'Reilly, Lois Miller, and Verne Luckow, W.H. Freeman and Co., New York, USA). Five plaque purified virus preparations were produced from the virus preparation. Sf9 insect cells growing in suspension were infected with each of the plaque purified virus preparations at a multiplicity of 0.5

chondroitinase ABC and keratinase (10u/ml) at 37C overnight. The samples were then examined by Western blotting using an antiserum that reacts with a neoepitope generated when aggrecan is cleaved by aggrecanase.

Another method for expression of mpts-10 was the drosophila expression system. The DNA fragment containing the sequences encoding mpts-10 and flanked by Not-1 and Sfi-1 that had been generated by PCR (see above) was cloned into plasmid Cmk 33. Cmk33 is a plasmid derived from pMK33/pMtHy (Li, Bin et al Biochem J (1996) 313, 57-64) so that Not-1 and Sfi-1 were in the cloning site. The overhangs generated by digestion of this plasmid are compatible with the overhangs generated in the digested DNA containing the mpts-10 fragment (see above). A plasmid containing the correct sequence of mpts-10 was amplified and purified. Drosophila (S2) cells were transformed with the plasmid using standard techniques (Li, Bin et al Biochem J (1996) 313, 57-64). Culture fluid was collected 2 days after transfection. These samples were assayed for aggrecanase activity by incubating with bovine aggrecan (Sigma, St. Louis, MO, USA) at a concentration of 0.1 mg/ml. The samples were then incubated with both chondroitinase ABC and keratinase (10u/ml) at 37C overnight. The samples were then examined by Western blotting using an antiserum that reacts with a neoepitope generated when aggrecan is cleaved by aggrecanase.

Example 3

Expression of MPTS-15

An example of a system for expression of mpts-15 is the COS-7 mammalian cell system. The nucleotide sequence that encodes mpts-15, including the secretion signal sequence, was ligated into a pcDNA3.1 plasmid (In Vitrogen, Carlsbad, CA, USA). Two micrograms of the resulting plasmid was combined with lipofectamine (Life Technologies, Rockville, MD, USA). The mixture was then added to COS-7 cells, which were grown in 6 well plates to a density of approximately 90% confluency. After 6 hours, fresh medium was added to the cells and after 24 hours the cells were washed and fresh serum free medium containing bovine aggrecan (0.1mg/ml, Sigma, St. Louis, MO, USA) was added. The cells incubated for an additional 48 hours. Five hundred

microliters of culture fluid from each well was collected and concentrated ten fold. Two microliters of chondroitinase ABC and keratinase (10 u/ml, Sigma, St. Louis, MO, USA) was then added and the samples incubated overnight at 37C. The samples were then boiled in SDS-PAGE sample loading buffer, electrophoresed on a polyacryamide gel and transferred to a PVDF membrane. A Western blot using an antiserum against a neoepitope generated when aggrecanase cleaves aggrecan was then performed.

Another example of a system for expression of mpts-15 was the baculovirus expression system. The DNA sequence that contained the coding sequence for mpts-15 (including the sequences that code for the secretion signal sequence) and that had been cloned in the pcDNA3.1 vector was modified by PCR so that the coding sequence and the translational stop codon were flanked by the Not 1 (N-terminal side) and Sfi-1 (C-terminal side). The primer used for the N-terminal end was GATCGCGGCCGCTATGGAAATTTTGTGGAAGACGTTG and the primer for the C-terminal end was TGAGGCCTTCAGGGCCGATCTTAAAGCAAAGTTTCTTTTGGT. After amplification using standard PCR methods, the fragment was digested with Not 1 and Sfi-1. The digested fragment was ligated into a vector pVL1392-U, which had also been digested with Not1 and Sfi-1. PVL1392-U is a derivation of the baculovirus transfer plasmid, pVL1392 (PharMingen, San Diego, CA, USA) in which the multiple cloning site has been modified to contain Not-1 and Sfi-1. The overhangs generated by digestion with Not-1 and Sf1-1 were complementary to the overhangs generated in the Not 1 and Sfi 1digested PCR amplified DNA. The ligated DNA was transformed into bacterial cells and a clone was selected that contained the plasmid and the correct mpts 15 sequence. This plasmid was produced and purified. The mpts-15 sequence was transferred into a baculovirus vector using standard techniques (*Baculovirus Expression Vectors: A Laboratory Manual* by David O'Reilly, Lois Miller, and Verne Luckow, W.H. Freeman and Co., New York, USA). Five plaque purified virus preparations were produced from the virus preparation. Sf9 insect cells growing in suspension were infected with each of the plaque purified virus preparations at a multiplicity of 0.5. Culture fluid was harvested at 3 days after infection. These samples were assayed for

examined by Western blotting using an antiserum that reacts with a neoepitope generated when aggrecan is cleaved by aggrecanase.

Another method for expression of mpts-15 was the drosophila expression system. The DNA fragment containing the sequences encoding mpts-15 and flanked by Not-1 and Sfi-1 that had been generated by PCR (see above) was cloned into plasmid Cmk 33. Cmk33 is a plasmid derived from pMK33/pMtHy (Li, Bin et al Biochem J (1996) 313, 57-64) so that Not-1 and Sfi-1 were in the cloning site. The overhangs generated by digestion of this plasmid are compatible with the overhangs generated in the Not 1 and Sfi 1 digested DNA containing the mpts-15 fragment (see above). A plasmid containing the correct sequence of mpts-15 was amplified and purified. Drosophila (S2) cells were transformed with the plasmid using standard techniques (Li, Bin et al Biochem J (1996) 313, 57-64). Culture fluid was collected 2 days after transfection. . These samples were assayed for aggrecanase activity by incubating with bovine aggrecan (Sigma,St. Louis, MD, USA) at a concentration of 0.1 mg/ml. The samples were then incubated with both chondroitinase ABC and keratinase (10u/ml) at 37C overnight. The samples were then examined by Western blotting using an antiserum that reacts with a neoepitope generated when aggrecan is cleaved by aggrecanase.

Example 4

Expression of MPTS-19

An example of a system for expression of mpts-19 is the COS-7 mammalian cell system. The nucleotide sequence that encodes mpts-19, including the secretion signal sequence and the C-terminal stop codon, was ligated into a pcDNA3.1 plasmid (In Vitrogen, Carlsbad, CA, USA). Two micrograms of the resulting plasmid was combined with lipofectamine (Life Technologies, Rocheville, MD, USA). The mixture was then added to COS-7 cells, which were grown in 6 well plates to a density of approximately 90% confluency. After 6 hours, fresh medium was added to the cells and after 24 hours the cells were washed and fresh serum free medium containing bovine aggrecan (0.1mg/ml, Sigma, St. Louis, MO, USA) was added. The cells incubated for an

additional 48 hours. Five hundred micoliters of culture fluid from each well was collected and concentrated ten fold. Two microliters of chondroitinase ABC and keratinase (10 u/ml, Sigma, St. Louis, MO, USA) was then added and the samples incubated overnight at 37C. The samples were then boiled in SDS-PAGE sample loading buffer, electrophoresed on a polyacryamide gel and transferred to a PVDF membrane. A Western blot using an antiserum against a neoepitope generated when aggrecanase cleaves aggrecan was then performed.

Another example of a system for expression of mpts-19 was the baculovirus expression system. The DNA sequence that contained the coding sequence for mpts-19 (including the sequences that code for the secretion signal sequence) and that had been cloned in the pcDNA3.1 vector was modified by PCR so that the coding sequence and the translational stop codon were flanked by the Not 1 (N-terminal side) and Sfi-1 (C-terminal side). The primer used for the N-terminal end was GATCGCGGCCGCTATGCCCGGCGGCCC CAGTCCCCG and the primer for the C-terminal end was TGAGGCCTTCAGGGCCGATCTCAGCGGCGGGCAACCCGCTG. After amplification using standard PCR methods, the fragment was digested with Not 1 and Sfi-1. The digested fragment was ligated into a vector pVL1392-U, which had also been digested with Not1 and Sfi-1. PVL1392-U is a derivation of the baculovirus transfer plasmid, pVL1392 (PharMingen, San Diego, CA USA) in which the multiple cloning site has been modified to contain Not-1 and Sfi-1. The overhangs generated by digestion with Not-1 and Sfi-1 were complementary to the overhangs generated in the Not 1 and Sfi 1 digested PCR amplified DNA. The ligated DNA was transformed into bacterial cells and a clone was selected that contained the plasmid and the correct mpts-19 sequence. This plasmid was produced and purified. The mpts-19 sequence was transferred into a baculovirus vector using standard techniques (*Baculovirus Expression Vectors: A Laboratory Manual* by David O'Reilly, Lois Miller, and Verne Luckow, W.H. Freeman and Co., New York, USA). Five plaque purified virus preparations were produced from the virus preparation. Sf9 insect cells growing in suspension were

[illegible] aggrecan (Sigma, St. Louis, MO, USA) at

a concentration of 0.1 mg/ml. The samples were then incubated with both chondroitinase ABC and keratinase (10u/ml) at 37C overnight. The samples were then examined by Western blotting using an antiserum that reacts with a neoepitope generated when aggrecan is cleaved by aggrecanase.

Another method for expression of mpts-19 was the drosophila expression system. The DNA fragment containing the sequences encoding mpts-19 and flanked by Not-1 and Sfi-1 that had been generated by PCR (see above) was cloned into plasmid Cmk 33. Cmk33 is a plasmid derived from pMK33/pMtHy (Li, Bin et al Biochem J (1996) 313, 57-64) so that Not-1 and Sfi-1 were in the cloning site. The overhangs generated by digestion of this plasmid with Not 1 and Sfi 1 are compatible with the overhangs generated in the digested DNA containing the mpts-19 fragment (see above). A plasmid containing the correct sequence of mpts-19 was amplified and purified. Drosophila (S2) cells were transformed with the plasmid using standard techniques (Li, Bin et al Biochem J (1996) 313, 57-64). Culture fluid was collected 2 days after transfection. These samples were assayed for aggrecanase activity by incubating with bovine aggrecan (Sigma, St. Louis, MO, USA) at a concentration of 0.1 mg/ml. The samples were then incubated with both chondroitinase ABC and keratinase (10u/ml) at 37C overnight. The samples were then examined by Western blotting using an antiserum that reacts with a neoepitope generated when aggrecan is cleaved by aggrecanase.

Example 5

Expression of MPTS-20

An example of a system for expression of mpts-20 is the COS-7 mammalian cell system. The nucleotide sequence that encodes mpts-10, including the secretion signal sequence and the C-terminal stop codon, was ligated into a pcDNA3.1 plasmid (In Vitrogen, Carlesbad, CA, USA). Two micrograms of the resulting plasmid was combined with lipofectamine (Life Technologies, Rockeville, MD, USA). The mixture was then added to COS-7 cells, which were grown in 6 well plates to a density of approximately 90% confluency. After 6 hours, fresh medium was added to the cells and

after 24 hours the cells were washed and fresh serum free medium containing bovine aggrecan (0.1mg/ml, Sigma, St. Louis, MO, USA) was added. The cells incubated for an additional 48 hours. Five hundred micoliters of culture fluid from each well was collected and concentrated ten fold. Two microliters of chondroitinase ABC and keratinase (10 u/ml, Sigma, St. Louis, MO, USA) was then added and the samples incubated overnight at 37C. The samples were then boiled in SDS-PAGE sample loading buffer, electrophoresed on a polyacryamide gel and transferred to a PVDF membrane. A Western blot using an antiserum against a neoepitope generated when aggrecanase cleaves aggrecan was then performed.

Another example of a system for expression of mpts-20 was the baculovirus expression system. The DNA sequence that contained the coding sequence for mpts-20 (including the sequences that code for the secretion signal sequence) and that had been cloned in the pcDNA3.1 vector was modified by PCR so that the coding sequence and the translational stop codon were flanked by the Not 1 (N-terminal side) and Sfi-1 (C-terminal side). The primer used for the N-terminal end was GATCGCGGCCGCTGCGCTGTGATGAGTGTGCCTG and the primer for the C-terminal end was TGAGGCCTTCAGGGCCGATCTTATAAAGGCCTTGAGAAAACAG. After amplification using standard PCR methods, the fragment was digested with Not 1 and Sfi-1. The digested fragment was ligated into a vector pVL1392-U, which had also been digested with Not1 and Sfi-1. PVL1392 U is a derivation of the baculovirus transfer plasmid, pVL1392 (PharMingen, San Diego, CA USA) in which the multiple cloning site has been modified to contain Not 1 and Sfi-1. The overhangs generated by digestion with Not 1 and Sfi-1 were complementary to the overhangs generated in the Not 1 and Sfi 1 digested PCR amplified DNA. The ligated DNA was transformed into bacterial cells and a clone was selected that contained the plasmid and the correct mpts-20 sequence. This plasmid was produced and purified. The mpts-20 sequence was transferred into a baculovirus vector using standard techniques (*Baculovirus Expression Vectors: A Laboratory Manual* [illegible]

[illegible] with each of the plaque purified virus preparations at a multiplicity of 1-5.

Culture fluid was harvest 3 days after infection. These samples were assayed for aggrecanase activity by incubating with bovine aggrecan (Sigma, St. Louis, MO, USA) at a concentration of 0.1 mg/ml. The samples were then incubated with both chondroitinase ABC and keratinase (10u/ml) at 37C overnight. The samples were then examined by Western blotting using an antiserum that reacts with a neoepitope generated when aggrecan is cleaved by aggrecanase.

Another method for expression of mpts-20 was the drosophila expression system. The DNA fragment containing the sequences encoding mpts-20 and flanked by Not-1 and Sfi-1 that had been generated by PCR (see above) was cloned into plasmid Cmk 33. Cmk33 is a plasmid derived from pMK33/pMtHy (Li, Bin et al Biochem J (1996) 313, 57-64) so that Not-1 and Sfi-1 were in the cloning site. The overhangs generated by digestion of this plasmid with Not 1 and Sfi 1 are compatible with the overhangs generated in the digested DNA containing the mpts-20 fragment (see above). A plasmid containing the correct sequence of mpts-20 was amplified and purified. Drosophila (S2) cells were transformed with the plasmid using standard techniques (Li, Bin et al Biochem J (1996) 313, 57-64). Culture fluid was collected 2 days after transfection. These samples were assayed for aggrecanase activity by incubating with bovine aggrecan (Sigma, St. Louis, MO, USA) at a concentration of 0.1 mg/ml. The samples were then incubated with both chondroitinase ABC and keratinase (10u/ml) at 37C overnight. The samples were then examined by Western blotting using an antiserum that reacts with a neoepitope generated when aggrecan is cleaved by aggrecanase.

Example 6

Purification of mpts-10, 15, 19 and 20:

Mpts-10, 15, 19, and 20 were purified from the culture fluid of the expression systems described above using chromatographic procedures. For example, the culture fluid was adjusted with regard to pH, filtered and then loaded onto a column packed with sulfopropyl sepharose FF (Amersham-Pharmacia Biotech, Piscataway, NJ, USA). After washing with a buffer consisting of 10 mM $CaCl_2$, 0.1 M NaCl, and 0.05% Brij35 at a pH

which results in retention of the mpts's on the column, the mpts's were eluted with a 0.1 M to 1.0 M NaCl gradient. Fractions from the column were assayed for the presence of aggrecanase activity as described above and noded. For the purification a column packed with phenylsepharose or sephacryl S-200 can also be used.

SEQUENCE LISTING

<110> Hoffmann-La Roche AG

<120> Novel Metalloproteases Having Thrombospondin Domains and Nucleic Acid Compositions Encoding the Same

<130> 20594

<150> 60/184,152

<151> 2000-02-18

<160> 10

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 959

<212> PRT

<213> human

<220>

<223> Xaa at 909 is any amino acid

<223> X at 211 is any amino acid

<223> X at 218 is any amino acid

<223> Xaa is any amino acid

<400> 1

Met Glu Ile Leu Trp Lys Thr Leu Thr Trp Ile Leu Ser Leu Ile Met
1 5 10 15

Ala Ser Ser Glu Phe His Ser Asp His Arg Leu Ser Tyr Ser Ser Gln
20 25 30

Glu Glu Phe Leu Thr Tyr Leu Glu His Tyr Gln Leu Thr Ile Pro Ile
35 40 45

Arg Val Asp Gln Asn Gly Ala Phe Leu Ser Phe Thr Val Lys Asn Asp
50 55 60

Lys His Ser Arg Arg Arg Arg Ser Met Asp Pro Ile Asp Pro Gln Gln
65 70 75 80

Ala Val Ser Lys Leu Phe Phe Lys Leu Ser Ala Tyr Gly Lys His Phe
85 90 95

His Leu Asn Leu Thr Leu Asn Thr Asp Phe Val Ser Lys His Phe Thr
100 105 110

Val Glu Tyr Trp Gly Lys Asp Gly Pro Gln Trp Lys His Asp Phe Leu
115 120 125

Asp Asn Cys His Tyr Thr Gly Tyr Leu Gln Asp Gln Arg Ser Thr Thr
130 135 140

Lys Val Ala Leu Ser Asn Cys Val Gly Leu His Gly Val Ile Ala Thr
145 150 155 160

Glu Asp Glu Glu Tyr Phe Ile Gln Pro Leu Lys Asn Thr Thr Glu Asp
165 170 175

Ser Lys His Phe Ser Tyr [illegible] Asn Gly His Pro His Val Ile Tyr Lys
180 185 190

Lys Ser Ala Leu [illegible] Arg His Leu Tyr Asp His Ser His Cys Gly
195 200 205

[illegible]

[illegible]

Arg Gln Lys Arg Ser Val Ser Ile Glu Arg Phe Val Gln Thr Leu Val
245 250 255

Val Ala Asp Lys Met Met Val Gly Tyr His Gly Arg Lys Asp Ile [illegible]
260 265 270

His Tyr [illegible]

Leu Thr Glu Asp Gln Pro Asn Leu Glu Ile Asn His His Ala Asp Lys
305 310 315 320
Ser Leu Asp Ser Phe Cys Lys Trp Gln Lys Ser Ile Leu Ser His Gln
325 330 335
Ser Asp Gly Asn Thr Ile Pro Glu Asn Gly Ile Ala His His Asp Asn
340 345 350
Ala Val Leu Ile Thr Arg Tyr Asp Ile Cys Thr Tyr Lys Asn Lys Pro
355 360 365
Cys Gly Thr Leu Gly Leu Ala Ser Val Ala Gly Met Cys Glu Pro Glu
370 375 380
Arg Ser Cys Ser Ile Asn Glu Asp Ile Gly Leu Gly Ser Ala Phe Thr
385 390 395 400
Ile Ala His Glu Ile Gly His Asn Phe Gly Met Asn His Asp Gly Ile
405 410 415
Gly Asn Ser Cys Gly Thr Lys Gly His Glu Ala Ala Lys Leu Met Ala
420 425 430
Ala His Ile Thr Ala Asn Thr Asn Pro Phe Ser Trp Ser Ala Cys Ser
435 440 445
Arg Asp Tyr Ile Thr Ser Phe Leu Asp Ser Gly Arg Gly Thr Cys Leu
450 455 460
Asp Asn Glu Pro Pro Lys Arg Asp Phe Leu Tyr Pro Ala Val Ala Pro
465 470 475 480
Gly Gln Val Tyr Asp Ala Asp Glu Gln Cys Arg Phe Gln Tyr Gly Ala
485 490 495
Thr Ser Arg Gln Cys Lys Tyr Gly Glu Val Cys Arg Glu Leu Trp Cys
500 505 510
Leu Ser Lys Ser Asn Arg Cys Val Thr Asn Ser Ile Pro Ala Ala Glu
515 520 525
Gly Thr Leu Cys Gln Thr Gly Asn Ile Glu Lys Gly Trp Cys Tyr Gln
530 535 540
Gly Asp Cys Val Pro Phe Gly Thr Trp Pro Gln Ser Ile Asp Gly Gly
545 550 555 560
Trp Gly Pro Trp Ser Leu Trp Gly Glu Cys Ser Arg Thr Cys Gly Gly
565 570 575
Gly Val Ser Ser Ser Leu Arg His Cys Asp Ser Pro Ala Pro Ser Gly
580 585 590
Gly Gly Lys Tyr Cys Leu Gly Glu Arg Lys Arg Tyr Arg Ser Cys Asn
595 600 605
Thr Asp Pro Cys Pro Leu Gly Ser Arg Asp Phe Arg Glu Lys Gln Cys
610 615 620

Ala Asp Phe Asp Asn Met Pro Phe Arg Gly Lys Tyr Tyr Asn Trp Lys
625 630 635 640

Pro Tyr Thr Gly Gly Gly Val Lys Pro Cys Ala Leu Asn Cys Leu Ala
645 650 655

Glu Gly Tyr Asn Phe Tyr Thr Glu Arg Ala Pro Ala Val Ile Asp Gly
660 665 670

Thr Gln Cys Asn Ala Asp Ser Leu Asp Ile Cys Ile Asn Gly Glu Cys
675 680 685

Lys His Val Gly Cys Asp Asn Ile Leu Gly Ser Asp Ala Arg Glu Asp
690 695 700

Arg Cys Arg Val Cys Gly Gly Asp Gly Ser Thr Cys Asp Ala Ile Glu
705 710 715 720

Gly Phe Phe Asn Asp Ser Leu Pro Arg Gly Gly Tyr Met Glu Val Val
725 730 735

Gln Ile Pro Arg Gly Ser Val His Ile Glu Val Arg Glu Val Ala Met
740 745 750

Ser Lys Asn Tyr Ile Ala Leu Lys Ser Glu Gly Asp Asp Tyr Tyr Ile
755 760 765

Asn Gly Ala Trp Thr Ile Asp Trp Pro Arg Lys Phe Asp Val Ala Gly
770 775 780

Thr Ala Phe His Tyr Lys Arg Pro Thr Asp Glu Pro Glu Ser Leu Glu
785 790 795 800

Ala Leu Gly Pro Thr Ser Glu Asn Leu Ile Val Met Val Leu Leu Gln
805 810 815

Glu Gln Asn Leu Gly Ile Arg Tyr Lys Phe Asn Val Pro Ile Thr Arg
820 825 830

Thr Gly Ser Gly Asp Asn Glu Val Gly Phe Thr Trp Asn His Gln Ser
835 840 845

Trp Ser Gln Cys Ser Ala Thr Tyr Ala Gly Gly Lys Met Ile Thr Arg
850 855 860

Gln Val Thr Gln Arg Ala Arg Trp Arg Thr Cys His Ile Leu Ser Tyr
865 870 875 880

Ala Leu Cys Leu Leu Lys Lys Leu Ile Gly Asn Ile Ser Cys Arg Phe
885 890 895

Ala Ser Ser Cys Asn Leu Pro Lys Glu Thr Leu Leu Xaa Leu Tyr Tyr
900 905 910

Tyr Leu Glu Gly Gly Leu Phe Ala Phe Arg Glu His Ile Leu Gly
945 950 955

<210> 2
<211> 2879
<212> DNA
<213> human

<400> 2
atggaaattt tgtggaagac gttgacctgg attttgagcc tcatcatggc ttcatcggaa
60
tttcatagtg accacaggct ttcatacagt tcttaagagg aattcctgac ttatcttgaa
120
cactaccagc taactattcc aataagggtt gatcaaaatg gagcatttct cagctttact
180
gtgaaaaatg ataaatactc aaggagaaga cggagtatgg accctattga tctacagcag
240
gcagtattta agttattttt taaactttca gcctatggca agcactttca tctaaacttg
300
acttcaaaca cagattttgt gtccaaacat tttacagtag aatattgggg gaaagatgga
360
ccccagtgga aacatgattt tttagacaac tgttattaca caggatattt gtaagatgaa
420
cgtatataaa ctaaagtggc tttaagcaac tgtgttgggt tggatggtgt tattgctaca
480
gaagatgaag agtattttat cgaacettta aagaatacca cagaggattt caagtatttt
540
agttatgaaa atggccaccc tcatgttatt tacaaaaagt ctgcccttca aaaacgacat
600
ctgtatgatc actctcattg tggggtttcg gatttcacaa gaagtggcaa actttggtgg
660
ctgaatgaca catccactgt ttcttattca ctaccaatta acaacacaca tatccaccac
720
agacagaaga gatcagtgag cattgaacgg tttgtggaga cattggtagt ggcagacaaa
780
atgatggtgg gctaccatgg ccgcaaagac attgaacatt acattttgag tgtgatgaat
840
attgttgcca aactttaccg tgattccagc ctaggaaacg ttgtgaatat tatagtggcc
900

cgcttaattg ttctcacaga agatcagcca aacttggaga taaaccacca tgcagacaag
960
tccctcgata gcttctgtaa atggcagaaa tccattctct cccaccaaag tgatggaaac
1020
accattccag aaaatgggat tgcccaccac gataatgcag ttcttattac tagatatgat
1080
atctgcactt ataaaaataa gccctgtgga acactgggct tggcctctgt ggctggaatg
1140
tgtgagcctg aaaggagctg cagcattaat gaagacattg gcctgggttc agcttttacc
1200
attgcacatg agattggtca caattttggt atgaaccatg atggaattgg aaattcttgt
1260
gggacgaaag gtcatgaagc agcaaaactt atggcagctc acattactgc gaataccaat
1320
cctttttcct ggtctgcttg cagtcgagac tacatcacca gctttctaga ttcaggccgt
1380
ggtacttgcc ttgataatga gcctcccaag cgtgactttc tttatccagc tgtggcccca
1440
ggtcaggtgt atgatgctga tgagcaatgt cgtttccagt atggagcaac ctcccgccaa
1500
tgtaaatatc gggaagtgtg tagagagctc tggtgtcgca gcaaaagcaa ccgctgtgtc
1560
accaacagta ttccagcagc tgagggcaca ctgtgtcaaa ctgggaatat tgaaaaaggg
1620
tggtgttatc aggagattg tgttctttt [illegible] [illegible] cctagagcat agatggggc
1680
[illegible]
1740
[illegible]
1800
[illegible]
1860
gagaaacagt gtgcagactt tgacaatatg cctttccgag gaaagtatta taactggaaa
1920
[illegible] gaggtgggt [illegible] [illegible] actggctga aggttataat
[illegible]

gctagggaag atagatgtcg agtctgtgga ggggacggaa gcacatgtga tgccattgaa 2160

gggttcttca atgattcact gcccagggga ggctacatgg aagtggtgca gataccaaga 2220

ggctctgttc acattgaagt tagagaagtt gccatgtcaa agaactatat tgctttaaaa 2280

tctgaaggag atgattacta tattaatggt gcctggacta ttgactggcc taggaaattt 2340

gatgttgctg ggacagcttt tcattacaag agaccaactg atgaaccaga atccttggaa 2400

gctctaggtc ctacctcaga aaatctcatc gtcatggttc tgcttcaaga acagaatttg 2460

ggaattaggt ataagttcaa tgttcccatc actcgaactg gcagtggaga taatgaagtt 2520

ggctttacat ggaatcatca gtcttggtca gaatgctcag ctacttgtgc tggaggtaag 2580

atgcccacta ggcagcccac ccagagggca agatggagaa caaaacacat tctgagctat 2640

gctttgtgtt tgttaaaaaa gctaattgga aacatttctt gcaggtttgc ttcaagctgt 2700

catttaccaa aagaaacttt gctttaatta tattatattt catttgtttt caacctcatg 2760

taatttgtgc agatttgttg gtaaaataca tcttggcaca atgagtgtct ttgctggtgc 2820

tctcccaag actatcttga aggtgggctg tttgcctttc atgaaaaaat tcttggtat 2879

<210> 3

<211> 947

<212> PRT

<213> human

<400> 3

Met Ala Pro Ala Cys Gln Ile Leu Arg Trp Ala Leu Ala Leu Gly Leu
1 5 10 15

Gly Leu Met Phe Glu Val Thr His Ala Phe Arg Ser Gln Asp Glu Phe
20 25 30

Leu Ser Ser Leu Glu Ser Tyr Glu Ile Ala Phe Pro Thr Arg Val Asp
35 40 45

His Asn Gly Ala Leu Leu Ala Phe Ser Pro Pro Pro Pro Arg Arg Gln
50 55 60
Arg Arg Gly Thr Gly Ala Thr Ala Glu Ser Arg Leu Phe Tyr Lys Val
65 70 75 80
Ala Ser Pro Ser Thr His Phe Leu Leu Asn Leu Thr Arg Ser Ser Arg
85 90 95
Leu Leu Ala Gly His Val Ser Val Glu Tyr Trp Thr Arg Glu Gly Leu
100 105 110
Ala Trp Gln Arg Ala Ala Arg Pro His Cys Leu Tyr Ala Gly His Leu
115 120 125
Gln Gly Gln Ala Ser Ser Ser His Val Ala Ile Ser Thr Cys Gly Gly
130 135 140
Leu His Gly Leu Ile Val Ala Asp Glu Glu Glu Tyr Leu Ile Glu Pro
145 150 155 160
Leu His Gly Gly Pro Lys Gly Ser Arg Ser Pro Glu Glu Ser Gly Pro
165 170 175
His Val Val Tyr Lys Arg Ser Ser Leu Arg His Pro His Leu Asp Thr
180 185 190
Ala Cys Gly Val Arg Asp Glu Lys Pro Trp Lys Gly Arg Pro Trp Trp
195 200 205
Leu Arg Thr Leu Lys Pro Pro Pro Ala Arg Pro Leu Gly Asn Glu Thr
210 215 220
Glu Arg Gly Gln Pro Gly Leu Lys Arg Ser Val Ser Arg Glu Arg Tyr
225 230 235 240
Val Glu Thr Leu Val Val Ala Asp Lys Met Met Val Ala Tyr His Gly
245 250 255
Arg Arg Asp Val Glu Ile Lys Val Leu Ala Ile Met Asn Ile Val Ala
260 265 270
Lys Val Thr Asp Ser Ser Leu Gly Ser Thr Val Asn Ile Leu Leu Val
275 280 285
Thr Arg Leu Leu Leu Leu Thr Glu Asp Gln Pro Thr Leu Glu Ile Thr
290 295 300
His His Ala Gly Lys Ser Leu Asp Ser Phe Cys Lys Trp Gln Lys Ser
305 310 315 320
Ile Val Asn His Ser Gly His Gly Asn Ala Ile Pro Glu Asn Gly Val
325 330 335

Met Cys Glu Arg Glu Arg Ser Cys Ser Val Asn Glu Asp Ile Gly Leu
370 375 380
Ala Thr Ala Phe Thr Ile Ala His Glu Ile Gly His Thr Phe Gly Met
385 390 395 400
Asn His Asp Gly Val Gly Asn Ser Cys Gly Ala Arg Gly Gln Asp Pro
405 410 415
Ala Lys Leu Met Ala Ala His Ile Thr Met Lys Thr Asn Pro Phe Val
420 425 430
Trp Ser Ser Cys Ser Arg Asp Tyr Ile Thr Ser Phe Leu Asp Ser Gly
435 440 445
Leu Gly Leu Cys Leu Asn Asn Arg Pro Pro Arg Gln Asp Phe Val Tyr
450 455 460
Pro Thr Val Ala Pro Gly Gln Ala Tyr Asp Ala Asp Glu Gln Cys Arg
465 470 475 480
Phe Gln His Gly Val Lys Ser Arg Gly Leu Gln Arg Ala Val Val Ser
485 490 495
Glu Gln Glu Gln Pro Val His His Gln Gln His Pro Gly Arg Arg Gly
500 505 510
His Ala Val Pro Asp Ala His His Arg Gln Gly Val Val Leu Gln Thr
515 520 525
Gly Leu Cys Pro Leu Trp Val Ala Pro Arg Gly Cys Gly Arg Ser Leu
530 535 540
Gly Ala Val Asp Ser Met Gly Asp Cys Ser Arg Thr Cys Gly Gly Gly
545 550 555 560
Val Ser Ser Ser Ser Arg His Cys Asp Ser Pro Arg Pro Thr Ile Gly
565 570 575
Gly Lys Tyr Cys Leu Gly Glu Arg Arg Arg His Arg Ser Cys Asn Thr
580 585 590
Asp Asp Cys Pro Pro Gly Ser Gln Asp Phe Arg Glu Val Gln Cys Ser
595 600 605
Glu Phe Asp Ser Ile Pro Phe Arg Gly Lys Phe Tyr Lys Trp Lys Thr
610 615 620
Tyr Arg Gly Gly Gly Val Lys Ala Cys Ser Leu Thr Cys Leu Ala Glu
625 630 635 640
Gly Phe Asn Phe Tyr Thr Glu Arg Ala Ala Ala Val Val Asp Gly Thr
645 650 655
Pro Cys Arg Pro Asp Thr Val Asp Ile Cys Val Ser Gly Glu Cys Lys
660 665 670
His Val Gly Cys Asp Arg Val Leu Gly Ser Asp Leu Arg Glu Asp Lys
675 680 685

Cys Arg Val Cys Gly Gly Asp Gly Ser Ala Cys Glu Thr Ile Glu Gly
690 695 700

Val Phe Ser Pro Ala Ser Pro Gly Ala Gly Tyr Glu Asp Val Val Trp
705 710 715 720

Ile Pro Lys Gly Ser Val His Ile Phe Ile Gln Asp Leu Asn Leu Ser
725 730 735

Leu Ser His Leu Ala Leu Lys Gly Asp Gln Glu Ser Leu Leu Leu Glu
740 745 750

Gly Leu Pro Gly Thr Pro Gln Pro His Arg Leu Pro Leu Ala Gly Thr
755 760 765

Thr Phe Gln Leu Arg Gln Gly Pro Asp Gln Val Gln Ser Leu Glu Ala
770 775 780

Leu Gly Pro Ile Asn Ala Ser Leu Ile Val Met Val Leu Ala Arg Thr
785 790 795 800

Glu Leu Pro Ala Leu Arg Tyr Arg Phe Asn Ala Pro Ile Ala Arg Asp
805 810 815

Ser Leu Pro Pro Tyr Ser Trp His Tyr Ala Pro Trp Thr Lys Cys Ser
820 825 830

Pro Ser Val Gln Ala Val Ala Arg Cys Arg Arg Trp Ser Ala Ala Thr
835 840 845

Lys Leu Asp Ser Ser Ala Val Ala Pro His Tyr Cys Ser Ala His Ser
850 855 860

Lys Leu Ala Gln Lys Gln Ala Arg Leu Gln His Gly Ala Leu Pro Gln
865 870 875 880

Asp Trp Val Val Gly Thr Val Ala Leu Gln Pro Gln Leu Ala Met Gln
885 890 895

Gly Val Arg Ser Arg Ser Val Val Cys Gln Ala Pro Arg Leu Cys Arg
[illegible] [illegible] [illegible]

Gln [illegible] Tyr Ala Leu Arg Asp Ser Ala Cys Ile [illegible] [illegible] Arg [illegible] [illegible]
[illegible] [illegible] [illegible]

[illegible] Leu Arg [illegible] Ala Thr Ala [illegible] Leu Ala [illegible] Arg [illegible] [illegible] Gly [illegible]
[illegible] 935 [illegible]

Arg Leu Val
945

<400> 4

ttccatccta atacgactca ctatagggct cgagcggccg cccgggcagg tgtggacacg 60

tggcctctat ggctcccgcc tgccagatcc tccgctgggc cctcgccctg gggctgggcc 120

tcatgttcga ggtcacgcat gccttccggt ctcaagatga gttcctgtcc agtctggaga 180

gctatgagat cgccttcccc acccgcgtgg accagaacgg ggcactgctg gccttctcgc 240

cacctcctcc ccggaggcag cgccgcggca cgggggccac agccgagtcc cgcctcttct 300

acaaagtggc ctcgcccagc acccacttcc tgctgaacct gacccgcagc tcccgtctac 360

tggcagggca cgtctccgtg gagtactgga cacgggaggg cctggcctgg cagagggcgg 420

cccggctcca ctgcctctac gctggtcacc tgcagggcca ggccagcagc tcccatgtgg 480

ccatcagcac ctgtggaggc ctgcacggcc tgatcgtggc agacgaggaa gagtacctga 540

ttgagccctt gcacggtggc cccaagggtt ctcggagccc ggaggaaagt ggaccacatc 600

tggtgtacaa gcgttcctct ctgcgtcacc cccacctgga cacagcctgt ggagtgagag 660

atgagaaacc gtggaaaggg cggccctggt ggctgcggac cttgaagcca ccgcctgcca 720

ggcccctggg gaatgaaaca gagcgtggcc agccaggcct gaagcgatcg gtcagccgag 780

agcgctacgt ggagaccctg gtggtcgctg acaagatgat ggtggcctat cacgggcgcc 840

gggatatgga gcagtatgtc ctggccatca tgaacattgt tgccaaactt ttccaggact 900

cgagtctggg aagcaccgtt aacatcctcg taactcgcct catcctgctc acggaggacc 960

agcccactct ggagatcacc caccatgccg ggaagtccct ggacagcttc tgtaagtggc 1020

agaaatccat cgtgaaccac agcggccatg gcaatgccat tccagagaac ggtgtggcta 1080

accatgacac agcagtgctc atcacacgct atgacatctg catctacaag aacaaaccct 1140

gcggcacact aggcctggcc ccggtgggcg gaatgtgtga gcgcgagaga agctgcagcg
1200
tcaatgagga cattggcctg gccacagcgt tcaccattgc ccacgagatc gggcacacat
1260
tcggcatgaa ccatgacggc gtgggaaaca gctgtggggc ccgtggtcag gacccagcca
1320
agctcatggc tgcccacatt accatgaaga ccaacccatt cgtgtggtca tcctgcagcc
1380
gtgactacat caccagcttt ctagactcgg gcctggggct ctgcctgaac aaccggcccc
1440
ccagacagga ctttgtgtac ccgacagtgg caccgggcca agcctacgat gcagatgagc
1500
aatgccgctt tcagcatgga gtcaaatcgc gaggtctgca gcgagctgtg gtgtctgagc
1560
aagagcaacc ggtgcatcac caacagcatc ccggccgccg agggcacgct gtgccagacg
1620
cacaccatcg acaaggggtg gtgctacaaa cgggtctgtg tcccctttgg gtcgcgccca
1680
gagggtgtgg acggagcctg ggggccgtgg actccatggg ggactgcagc cggacctgtg
1740
gcggcggcgt gtcctcttct agccgtcact gcgacagccc caggccaacc atcgggggca
1800
agtactgtct gggtgagaga aggcggcacc gcttctgcaa cacggatgac tgtccccctg
1860
gctcccagga cttcagagaa gtgcagtgtt c[illegible] cag[illegible]ccct ttccgtggga
1920
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
2100
gagtcctggg ctccgac[illegible]g cgggaggaca agt[illegible]ccgact gtgtgggggt gacggcagtg
2160
[illegible]
2220

cccagcccca ccgtctgcct ctagctggga ccacctttca actgcgacag gggccagacc
2400
aggttcagag cctcgaagcc ctgggaccga ttaatgcatc tctcatcgtc atggtgctgg
2460
cccggaccga gctgcctgcc ctccgctacc gcttcaatgc ccccatcgcc cgtgactcgc
2520
tgccctccta ctcctggcac tatgcgccct ggaccaagtg ctcgcccagt gtgcaggcgg
2580
tagccaggtg caggcggtgg agtgccgcaa ccaagctgga cagctccgcg gtcgcccccc
2640
actactgcag tgcccacagc aagcttgccc aaaagcaagc gcgcctgcaa catggagcct
2700
tgccccaaga ctgggttgta ggaactgtcg ctctgcagcc gcagcttgcg atgcaaggcg
2760
tgccctgccg ctcggtcgtg tgccaagcgc cgcgtctctg ccgcgaagaa aaggcgctgg
2820
acgactgcgc atgcccgcag ccgcgcccac ctgtactgag gcctgccacg gcccccactg
2880
ccctctggag tggcggccct cgactggtct gagtgcaccc ctagctgcgg gccgggcctc
2940
cgccaccgcg tcgtcccttg caagagcgca gaccaccgcg ccacgctgcc cccggcgcac
3000
tgctcacccg ccgccaagcc accggccacc atgcgctgca acttgcgccc ctgccccccg
3060
gcccgctggg tggctggcga gtggggtgag tgctctgcac agtgcggcct cgggcagcgg
3120
cagcgttcgg tgcgctgcac cagccacacg ggccaggcgt cgcacgagtg cacggaggcc
3180
ctgcggccgc cgactagtaa gcttcgtcga cccgggaatt aattccggac cggtacctgc
3240
aggcgtacca gctttcccta tag
3263

<210> 5
<211> 1690
<212> PRT
<213> human

<400> 5

Pro Val Pro Ala Met Pro Gly Gly Pro Ser Pro Arg Ser Pro Ala Pro
1 5 10 15

Leu Leu Arg Pro Leu Leu Leu Leu Leu Cys Ala Leu Ala Pro Gly Ala
20 25 30

Pro Gly Pro Ala Pro Gly Arg Ala Thr Glu Gly Arg Ala Ala Leu Asp
35 40 45

Ile Val His Pro Val Arg Val Asp Ala Gly Gly Ser Phe Leu Ser Tyr
50 55 60

Glu Leu Trp Pro Arg Ala Leu Arg Lys Arg Asp Val Ser Val Arg Arg
65 70 75 80

Asp Ala Pro Ala Phe Tyr Glu Leu Gln Tyr Arg Gly Arg Glu Leu Arg
85 90 95

Phe Asn Leu Thr Ala Asn Gln His Leu Leu Ala Pro Gly Phe Val Ser
100 105 110

Glu Thr Arg Arg Arg Gly Gly Leu Gly Arg Ala His Ile Arg Ala His
115 120 125

Thr Pro Ala Cys His Leu Leu Gly Glu Val Gln Asp Pro Glu Leu Glu
130 135 140

Gly Gly Leu Ala Ala Ile Ser Ala Cys Asp Gly Leu Lys Gly Val Phe
145 150 155 160

Gln Leu Ser Asn Glu Asp Tyr Phe Ile Glu Pro Leu Asp Ser Ala Pro
165 170 175

Ala Arg Pro Gly His Ala Gln Pro His Val Val Tyr Lys Arg Gln Ala
180 185 190

Pro Gln Arg Leu Ala Gln Arg Gly Asp Ser Ser Ala Pro Ser Thr Cys
195 200 205

Gly Val [illegible] Val [illegible] [illegible] Gln Leu [illegible] [illegible] Asn Arg [illegible] Arg Trp [illegible]
210 215 220

Gln Arg [illegible] [illegible] Trp Arg Arg Pro Arg Leu Arg Arg [illegible] [illegible] Gln [illegible]
225 230 235 240

Ser Val Ser Lys Glu Lys Trp Val Glu Thr Leu Val Val Ala Asp Ala
245 250 255

Lys Met Val Glu Tyr His Gly Gln Pro Gln Val Glu Ser Tyr Val Leu
260 265 270

Glu Glu Asp Leu Lys Ile Thr His His Ala Asp Asn Thr Pro Lys Ser
305 310 315 320
Phe Cys Lys Trp Gln Lys Ser Ile Asn Met Lys Gly Asp Ala His Pro
325 330 335
Leu His His Asp Thr Ala Ile Leu Leu Thr Arg Lys Asp Leu Cys Ala
340 345 350
Thr Met Asn Arg Pro Cys Glu Thr Leu Gly Leu Ser His Val Ala Gly
355 360 365
Met Cys Gln Pro His Arg Ser Cys Ser Ile Asn Glu Asp Thr Gly Leu
370 375 380
Pro Leu Ala Phe Thr Val Ala His Glu Leu Gly His Ser Phe Gly Ile
385 390 395 400
Gln His Asp Gly Ser Gly Asn Asp Cys Glu Pro Val Gly Lys Arg Pro
405 410 415
Phe Ile Met Ser Pro Gln Leu Leu Tyr Asp Ala Ala Pro Leu Thr Trp
420 425 430
Ser Arg Cys Ser Arg Gln Tyr Ile Thr Arg Phe Leu Asp Arg Gly Trp
435 440 445
Gly Leu Cys Leu Asp Asp Pro Pro Ala Lys Asp Ile Ile Asp Phe Pro
450 455 460
Met Val Pro Pro Gly Val Leu Tyr Asp Val Ser His Gln Cys Arg Leu
465 470 475 480
Gln Tyr Gly Ala Tyr Ser Ala Phe Cys Glu Asp Met Asp Asn Val Cys
485 490 495
His Thr Leu Trp Cys Ser Val Gly Thr Thr Cys His Ser Lys Leu Asp
500 505 510
Ala Ala Val Asp Gly Thr Arg Cys Gly Glu Asn Lys Trp Cys Leu Ser
515 520 525
Gly Glu Cys Val Pro Val Gly Phe Arg Pro Glu Ala Val Asp Gly Gly
530 535 540
Trp Ser Gly Trp Ser Ala Trp Ser Ile Cys Ser Arg Ser Cys Gly Met
545 550 555 560
Gly Val Gln Ser Ala Glu Arg Gln Cys Thr Gln Pro Thr Pro Lys Tyr
565 570 575
Lys Gly Arg Tyr Cys Val Gly Glu Arg Lys Arg Phe Arg Leu Cys Asn
580 585 590
Leu Gln Ala Cys Pro Ala Gly Arg Pro Ser Phe Arg His Val Gln Cys
595 600 605
Ser His Phe Asp Ala Met Leu Tyr Lys Gly Arg Leu His Thr Trp Val
610 615 620

```
Pro Val Val Asn Asp Val Asn Pro Cys Glu Leu His Cys Arg Pro Ala
625                 630                 635                 640
Asn Glu Tyr Phe Ala Glu Lys Leu Arg Asp Ala Val Val Asp Gly Thr
                645                 650                 655
Pro Cys Tyr Gln Val Arg Ala Ser Arg Asp Leu Cys Ile Asn Gly Ile
            660                 665                 670
Cys Lys Asn Val Gly Cys Asp Phe Glu Ile Asp Ser Gly Ala Met Glu
        675                 680                 685
Asp Arg Cys Gly Val Cys His Gly Asn Gly Ser Thr Cys His Thr Val
    690                 695                 700
Ser Gly Thr Phe Glu Glu Ala Glu Gly Leu Gly Tyr Val Asp Val Gly
705                 710                 715                 720
Leu Ile Pro Ala Gly Ala Arg Glu Ile Arg Ile Gln Glu Val Ala Glu
                725                 730                 735
Ala Ala Asn Phe Leu Ala Leu Arg Ser Glu Asp Pro Glu Lys Tyr Phe
            740                 745                 750
Leu Asn Gly Gly Trp Thr Ile Gln Trp Asn Gly Asp Tyr Gln Val Ala
        755                 760                 765
Gly Thr Thr Phe Thr Tyr Ala Arg Arg Gly Asn Trp Glu Asn Leu Thr
    770                 775                 780
Ser Pro Gly Pro Thr Lys Glu Pro Val Trp Ile Gln Leu Leu Phe Gln
785                 790                 795                 800
Glu Ser Asn Pro Gly Val His Tyr Glu Tyr Thr Ile His Arg Glu Ala
                805                 810                 815
Gly Gly His Asp Glu Val Pro Pro Pro Val Phe Ser Trp His Tyr Gly
            820                 825                 830
Pro Trp Thr Lys Cys Thr Val Thr Tyr Gly Arg Gly Val Gln Arg Gln
        [illegible]         [illegible]         [illegible]
Asn Val Tyr Cys Leu [illegible] Arg Gln Ala Gly Pro Val Asp Glu [illegible] His
    [illegible]         [illegible]         [illegible]
Cys Asp Pro Leu Gly Arg Ile Asp Asp Gln Gln Arg Lys Cys Ser Glu
865                 [illegible]         [illegible]         [illegible]
Gln Pro Cys Pro Ala Arg Trp Trp Ala Gly Glu Trp Gln Leu Cys Ser
                885                 [illegible]         895
Ser Ser Cys Gly Pro Gly Gly Leu Ser Arg Arg Ala Val Leu Cys Ile
            [illegible]                             [illegible]
```

Pro Cys Pro Ala Thr Trp Ala Val Gly Asn Trp Ser Gln Cys Ser Val
945 950 955 960
Thr Cys Gly Glu Gly Thr Gln Arg Arg Asn Val Leu Cys Thr Asn Asp
965 970 975
Thr Gly Val Pro Cys Asp Glu Ala Gln Gln Pro Ala Ser Glu Val Thr
980 985 990
Cys Ser Leu Pro Leu Cys Arg Trp Pro Leu Gly Thr Leu Gly Pro Glu
995 1000 1005
Gly Ser Gly Ser Gly Ser Ser Ser His Glu Leu Phe Asn Glu Ala Asp
1010 1015 1020
Phe Ile Pro His His Leu Ala Pro Arg Pro Ser Pro Ala Ser Ser Pro
1025 1030 1035 1040
Lys Pro Gly Thr Met Gly Asn Ala Ile Glu Glu Glu Ala Pro Glu Leu
1045 1050 1055
Asp Leu Pro Gly Pro Val Phe Val Asp Asp Phe Tyr Tyr Asp Tyr Asn
1060 1065 1070
Phe Ile Asn Phe His Glu Asp Leu Ser Tyr Gly Pro Ser Glu Glu Pro
1075 1080 1085
Asp Leu Asp Leu Ala Gly Thr Gly Asp Arg Thr Pro Pro Pro His Ser
1090 1095 1100
Arg Pro Ala Ala Pro Ser Thr Gly Ser Pro Val Pro Ala Thr Glu Pro
1105 1110 1115 1120
Pro Ala Ala Lys Glu Glu Gly Val Leu Gly Pro Trp Ser Pro Ser Pro
1125 1130 1135
Trp Pro Ser Gln Ala Gly Arg Ser Pro Pro Pro Pro Pro Ser Glu Gln Thr
1140 1145 1150
Pro Gly Asn Pro Leu Ile Asn Phe Leu Pro Glu Glu Asp Thr Pro Ile
1155 1160 1165
Gly Ala Pro Asp Leu Gly Leu Pro Ser Leu Ser Trp Pro Arg Val Ser
1170 1175 1180
Thr Asp Gly Leu Gln Thr Pro Ala Thr Pro Glu Ser Gln Asn Asp Phe
1185 1190 1195 1200
Pro Val Gly Lys Asp Ser Gln Ser Gln Leu Pro Pro Pro Trp Arg Asp
1205 1210 1215
Arg Thr Asn Glu Val Phe Lys Asp Asp Glu Glu Pro Lys Gly Arg Gly
1220 1225 1230
Ala Pro His Leu Pro Pro Arg Pro Ser Ser Thr Leu Pro Pro Leu Ser
1235 1240 1245
Pro Val Gly Ser Thr His Ser Ser Pro Ser Pro Asp Val Ala Glu Leu
1250 1255 1260

Trp Thr Gly Gly Thr Val Ala Trp Glu Pro Ala Leu Glu Gly Gly Leu
1265 1270 1275 1280
Gly Pro Val Asp Ser Glu Leu Trp Pro Thr Val Gly Val Ala Ser Leu
1285 1290 1295
Leu Pro Pro Pro Ile Ala Pro Leu Pro Glu Met Lys Val Arg Asp Ser
1300 1305 1310
Ser Leu Glu Pro Gly Thr Pro Ser Phe Pro Ala Pro Gly Pro Gly Ser
1315 1320 1325
Trp Asp Leu Gln Thr Val Ala Val Trp Gly Thr Phe Leu Pro Thr Thr
1330 1335 1340
Leu Thr Gly Leu Gly His Met Pro Glu Pro Ala Leu Asn Pro Gly Pro
1345 1350 1355 1360
Lys Gly Gln Pro Glu Ser Leu Ser Pro Glu Val Pro Leu Ser Ser Arg
1365 1370 1375
Leu Leu Ser Thr Pro Ala Trp Asp Ser Pro Ala Asn Ser His Arg Val
1380 1385 1390
Pro Glu Thr Gln Pro Leu Ala Pro Ser Leu Ala Glu Ala Gly Pro Pro
1395 1400 1405
Ala Asp Pro Leu Val Val Arg Asn Ala Ser Trp Gln Ala Gly Asn Trp
1410 1415 1420
Ser Glu Cys Ser Thr Thr Cys Gly Leu Gly Ala Val Trp Arg Pro Val
1425 1430 1435 1440
Arg Cys Ser Ser Gly Arg Asp Gln Asp Cys Ala Pro Ala Gly Arg Pro
1445 1450 1455
Gln Pro Ala Arg Arg Cys His Leu Arg Pro Cys Ala Thr Trp His Ser
1460 1465 1470
Lys Asn Trp Ser Cys Cys Ser Arg Ser Cys Gly Gly Gly Ser Gln Val
1475 1480 1485
Arg Asp Val His Cys Val Asp Thr Arg Asp Leu Arg Pro Leu Arg Ile
1490 1495 1500
Ile His Lys Gln Pro Gly Pro Ala Lys Pro Pro Ala His Asn Pro Cys
1505 1510 1515 1520
Gly Ala Gln Pro Cys Leu Ser Trp Tyr Thr Ser Ser Trp Arg Glu Cys
1525 1530 1535
Ser Gln Ala Cys Gly Gly Gly Glu Gln Gln Arg Leu Val Thr Cys Pro
1540 1545 1550

Cys Ser Ala Pro Cys Gly Gly Gly Val Gln Arg Arg Leu Val Lys Cys
1585 1590 1595 1600
Val Asn Thr Gln Thr Gly Leu Pro Glu Glu Asp Ser Asp Gln Cys Gly
1605 1610 1615
His Glu Ala Trp Pro Glu Ser Ser Arg Pro Cys Gly Thr Glu Asp Cys
1620 1625 1630
Glu Pro Val Glu Pro Pro Arg Cys Glu Arg Asp Arg Leu Ser Phe Gly
1635 1640 1645
Phe Cys Glu Thr Leu Arg Leu Leu Gly Arg Cys Gln Leu Pro Thr Ile
1650 1655 1660
Arg Thr Gln Cys Cys Arg Ser Cys Ser Pro Pro Ser His Gly Ala Pro
1665 1670 1675 1680
Ser Arg Gly His Gln Arg Val Ala Arg Arg
1685 1690

<210> 6
<211> 5338
<212> DNA
<213> human

<400> 6
ccggttcctg ccatgcccgg gggccccagt ccccgcagcc ccgcgccttt gctgcgcccc
60
ctcctcctgc tcctctgcgc tctggctccc ggcgcccccg gacccgcacc aggacgtgca
120
ccgagggccc gggcggcatt ggacatcgtg cacccggttc gagtcgacgc gggggctccc
180
ttcctgtcct acgagctgtg gctccgcgca ctgcgcaagc gggatgtatc tgtgcgccga
240
cacgcgcccg ccttctacca actacaatac cgcgggcgcg agctgcgctt caacctgacc
300
gccaatcagc acctgctggc gtccggcttt gtgagcgaga cgcggcggcg cggcggcctg
360
ggccgcgcgc acatccgggc ccacaccccg gcctgccacc tgcttggcga ggtgcaggac
420
cctgagctcg agggtggcct ggcggccatc agcgcctgcg acggcctgaa aggtgtgttc
480
cagctctcca acgaggacta cttcattgag cccctggaca gtgccccggc ccggcctggc
540

cacgcccagc cccatgtggt gtacaagcgt caggccccgg agaggctggc acagcggggt
600
gattccagtg ctccaagcac ctgtggagtg caagtgtacc cagagctgga gcctcgacgg
660
gagcgttggg agcagcggca gcagtggcgg cggccacggc tgaggcgtct acaccagcgg
720
tcggtcagca aagagaagtg ggtggagacc ctggtggtag ctgatgccaa aatggtggag
780
taccacggac agccgcaggt tgagagctat gtgctgacca tcatgaacat ggtggctggc
840
ctgtttcatg accccagcat tgggaacccc atccacatca ccattgtgcg cctggtcctg
900
ctggaagatg aggaggagga cctaaagatc acgcaccatg cagacaacac cccgaagagc
960
ttctgcaagt ggcagaaaag catcaacatg aagggggatg cccatcccct gcaccatgac
1020
actgccatcc tgctcaccag aaaggacctg tgtgcaacca tgaaccggcc ctgtgagacc
1080
ctgggactgt cccatgtggc gggcatgtgc cagccgcacc gcagctgcag catcaacgag
1140
gacacgggcc taccgctggc cttcac[illegible]gta g[illegible]ccac[illegible]agc [illegible]gggcacag ttttggcat[illegible]
1200
[illegible]agcatgacg [illegible]aagcggcaa [illegible]gac[illegible]gtg[illegible]g [illegible]ccg[illegible] [illegible]acgaccttt catcatgtct
1260
[illegible]acagctcc [illegible]gtacgacgc cg[illegible] [illegible] [illegible] [illegible]gcagccg ccagtatatc
1320
[illegible] [illegible] [illegible] [illegible] [illegible] [illegible]
[illegible]
[illegible] [illegible] [illegible] [illegible] [illegible] [illegible]
[illegible]
[illegible]agtacgg[illegible] [illegible]tactctgc cttctgcaa[illegible] [illegible] [illegible] [illegible] cacactctga
1500
[illegible]gctctg[illegible]gg ggaccacctg tcactccaag ct[illegible]gatg[illegible] [illegible]cgtggacgg cacccggtgt
1560
[illegible] [illegible] cag[illegible] [illegible] [illegible] [illegible]
[illegible]

tgtgtgggtg agcgcaagcg cttccgcctc tgcaacctgc aggcctgccc tgctggccgc
1800
ccctccttcc gccacgtcca gtgcagccac tttgatgcca tgctctacaa gggccggctg
1860
cacacatggg tgcccgtggt caatgacgtg aacccctgcg agctgcactg ccggcccgcg
1920
aatgagtact ttgccgagaa gctgcgggac gccgtggtcg atggcacccc ctgctaccag
1980
gtccgagcca gcagggacct ctgcatcaac ggcatctgta agaacgtggg ctgtgacttc
2040
gagattgact ccggtgctat ggaggatcgc tgtggtgtgt gccatggcaa cggctccacc
2100
tgccacaccg tgagcgggac ctttgaggag gccgagggcc tggggtatgt ggatgtgggg
2160
ctgatcccag cgggcgcatg cgagatccgc atccaagagg ttgcggaggc tgccaacttc
2220
ctggcactgc ggagcgagga cccggagaag tactttctca atggtggctg gaccatccag
2280
tggaacgggg actaccaggt ggcagggacc accttcacat acgcacgcag gggcaactgg
2340
gagaacctca cgtcccgggg tcctactaag gagcctgtct ggatccagct gctgttccag
2400
gagagcaacc ctggggtgca ctacgagtac accatccaca gggaggcagg tggccacgac
2460
gaggttctgc cgcccgtgtt ctcctggcat tatgggccct ggaccaagtg catagtcacc
2520
tgctggcgag gtgtgcagag gcagaatgtg tactgcttgg agcgggcagg agggcccgtg
2580
gacgaggagc actgtgaccc cctgggccgg cctgatgacc aacagaggaa gtgcagcgag
2640
cagccctgcc ctgccaggtg gtgggcaggt gagtggcagc tgtgctccag ctcctgcggg
2700
cctgggggcc tctcccgccg ggccgtgctc tgcatccgca gcgtggggct ggatgagcag
2760
aggggcctgg agccacccgc ctgtgaacac cttccccggc cccctactga aaccccttgc
2820
aaccgccatg taccctgtcc ggccacctgg gctgtgggga actggtctca gtgctcagtg
2880
acatgtgggg agggcactca gcgccgaaat gtcctctgca ccaatgacac cggtgtcccc
2940

tgtgacgagg cccagcagcc agccagcgaa gtcacctgct ctctgccact ctgtcggtgg
3000
cccctgggca cactgggccc tgaaggctca ggcagcggct cctccagcca cgagctcttc
3060
aacgaggctg acttcatccc gcaccacctg gccccacgcc cttcacccgc ctcatcaccc
3120
aagccaggca ccatgggcaa cgccattgag gaggaggctc cagagctgga cctgccgggg
3180
cccgtgtttg tggacgactt ctactacgac tacaatttca tcaatttcca cgaggatctg
3240
tcctacgggc cctctgagga gcccgatcta gacctggcgg ggacagggga ccggacaccc
3300
ccacccacca gccgtcctgc tgcgccctcc acgggtagcc ctgtgcctgc cacagagcct
3360
cctgcagcca aggaggaggg ggtactggga ccttggtccc cgagcccttg gcctagccag
3420
gccggccgct ccccaccccc accctcagag cagacccctg ggaacccttt gatcaatttc
3480
ctgcccgagg aagacacccc catagggccc ccagatcttg ggctccccag cctgtcctgg
3540
cccagggttt ccactgatgg cctgcagaca cctgccaccc ctgagagcca aaatgatttc
3600
ccagttggca aggacagcca gagccagctc ccccctccat ggcggga[illegible]ag gaccaatgag
3660
g[illegible]aagg atgatgagga [illegible]caaggac cgcggag[illegible]ac cccac[illegible] ccc[illegible]agacc[illegible]
3720
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
3900
at[illegible]ccctc t[illegible]ccagac[illegible] g[illegible]ggtcagg[illegible] ga[illegible]agttccc tggag[illegible]cggg ga[illegible]ccctc
3960
tt[illegible] [illegible] [illegible] [illegible]agactg tggca[illegible]gt[illegible] [illegible]
[illegible]

ccagcttggg acagccccgc caacagccac agagtccctg agacccagcc gctggctccc
4200
agcctggctg aagcggggcc ccccgcggac ccgttggttg tcaggaacgc cagctggcaa
4260
gcgggaaact ggagcgagtg ctctaccacc tgtggcctgg gtgcggtctg gaggccggtg
4320
cgctgtagct ccggccggga tgaggactgc gcccccgctg gccggcccca gcctgcccgc
4380
cgctgccacc tgcggccctg tgccacctgg cactcaggca actggagtaa gtgctcccgc
4440
agctgcggcg gaggttcctc agtgcgggac gtgcagtgtg tggacacacg ggacctccgg
4500
ccactgcggc ccttccattg tcagcccggg cctgccaagc cgcctgcgca ccggccctgc
4560
ggggcccagc cctgcctcag ctggtacaca tcttcctgga gggagtgctc cgaggcctgt
4620
ggcggtggtg agcagcagcg tctagtgacc tgcccggagc caggcctctg cgaggaggcg
4680
ctgagaccca acacctaccg gccctgcaac acccacccct gcacgcagtg ggtggtgggg
4740
cctgggggcc agtgctcagc cccctgtggt ggtggtgtcc agcgggcgct ggtcaagtgt
4800
gtcaacaccc agacagggct gcccgaggaa gacagtgacc agtgtggcca cgaggcctgg
4860
cctgcgagct cctggccgtg tggcactgac gattgtgagc ccgtcgagcc tcccgctgt
4920
gagccgaacc gcctgtcctt cgggttctgc gagacgctgc gcctactggg ccgctgccag
4980
ctgcctacca tccgcaccca gtgctgccgc tcgtgctctc cgcccagcca cggcgccccc
5040
tccccgggcc atcagcgggt tgcccgccgc tgactgtgcc aggatgcaca gaccgaccga
5100
cagacctcag tgcccaccac gggctgtggc ggagctcccg cccctgcgc cctaatggtg
5160
ctaacccccct ctcactaccc agcagcaggc tggggacctc ctccccctca aaaaaggtat
5220
ttttttattc taacagtttg tgtaacattt attatgattt tacataaatg agcatctacc
5280
attccaaagc aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaa
5338

<210> 7

<211> 947

<212> PRT

<213> human

<400> 7

Met Gln Phe Val Ser Trp Ala Thr Leu Leu Thr Leu Leu Val Arg Asp
1 5 10 15

Leu Ala Glu Met Gly Ser Pro Asp Ala Ala Ala Ala Val Arg Lys Asp
20 25 30

Arg Leu His Pro Arg Gln Val Lys Leu Leu Glu Thr Leu Ser Glu Tyr
35 40 45

Glu Ile Val Ser Pro Ile Arg Val Asn Ala Leu Gly Glu Pro Phe Pro
50 55 60

Thr Asn Val His Phe Lys Arg Thr Arg Arg Ser Ile Asn Ser Ala Thr
65 70 75 80

Asp Pro Trp Pro Ala Phe Ala Ser Ser Ser Ser Ser Ser Thr Ser Ser
85 90 95

Gln Ala His Tyr Arg Leu Ser Ala Phe Gly Gln Gln Phe Leu Phe Asn
100 105 110

Leu Thr Ala Asn Ala Gly Phe Ile Ala Pro Leu Phe Thr Val Thr Leu
115 120 125

Leu Gly Thr Pro Gly Val Asn Gln Thr Lys Phe Tyr Ser Glu Glu Gln
130 135 140

Ala Gln Leu Lys His [illegible] Phe [illegible] [illegible] Gly [illegible] [illegible] [illegible] [illegible] Asn [illegible]
145 150 155 160

[illegible]
[illegible]

Arg Ser His Asp Gly Asp Tyr Phe Ile Glu Pro Leu [illegible] Ser Met Asp
180 185 190

Gln Gln Glu Asp Glu Glu Glu Gln Asn Lys Pro His Ile Ile Tyr Arg
195 200 205

Arg Ser Ala Pro Gln Arg Glu Pro Ser Thr Gly Arg His Ala Cys Asp

[illegible]

Leu Asn Ser Gly Leu Ala Thr Glu Ala Phe Ser Ala Tyr Gly Asn Lys
260 265 270
Thr Asp Asn Thr Arg Glu Lys Arg Thr His Arg Arg Thr Lys Arg Phe
275 280 285
Leu Ser Tyr Pro Arg Phe Val Glu Val Leu Val Val Ala Asp Asn Arg
290 295 300
Met Val Ser Tyr His Gly Glu Asn Leu Gln His Tyr Ile Leu Thr Leu
305 310 315 320
Met Ser Ile Val Ala Ser Ile Tyr Lys Asp Pro Ser Ile Gly Asn Leu
325 330 335
Ile Asn Ile Val Ile Val Asn Leu Ile Val Ile His Asn Glu Gln Asp
340 345 350
Gly Pro Ser Ile Ser Phe Asn Ala Gln Thr Thr Leu Lys Asn Phe Cys
355 360 365
Gln Trp Gln His Ser Lys Asn Ser Pro Gly Gly Ile His His Asp Thr
370 375 380
Ala Val Leu Leu Thr Arg Gln Asp Ile Cys Arg Ala His Asp Lys Cys
385 390 395 400
Asp Thr Leu Gly Leu Ala Glu Leu Gly Thr Ile Cys Asp Pro Tyr Arg
405 410 415
Ser Cys Ser Ile Ser Glu Asp Ser Gly Leu Ser Thr Ala Phe Thr Ile
420 425 430
Ala His Glu Leu Gly His Val Phe Asn Met Pro His Asp Asp Asn Asn
435 440 445
Lys Cys Lys Glu Glu Gly Val Lys Ser Pro Gln His Val Met Ala Pro
450 455 460
Thr Leu Asn Phe Tyr Thr Asn Pro Trp Met Trp Ser Lys Cys Ser Arg
465 470 475 480
Lys Tyr Ile Thr Glu Phe Leu Asp Thr Gly Tyr Gly Glu Cys Leu Leu
485 490 495
Asn Glu Pro Glu Ser Arg Pro Tyr Pro Leu Pro Val Gln Leu Pro Gly
500 505 510
Ile Leu Tyr Asn Val Asn Lys Gln Cys Glu Leu Ile Phe Gly Pro Gly
515 520 525
Ser Gln Val Cys Pro Tyr Met Met Gln Cys Arg Arg Leu Trp Cys Asn
530 535 540
Asn Val Asn Gly Val His Lys Gly Cys Arg Thr Gln His Thr Pro Trp
545 550 555 560
Ala Asp Gly Thr Glu Cys Glu Pro Gly Lys His Cys Lys Tyr Gly Phe
565 570 575

Cys Val Pro Lys Glu Met Asp Val Pro Val Thr Asp Gly Ser Trp Gly
580 585 590

Ser Trp Ser Pro Phe Gly Thr Cys Ser Arg Thr Cys Gly Gly Gly Ile
595 600 605

Lys Thr Ala Ile Arg Glu Cys Asn Arg Pro Glu Pro Lys Asn Gly Gly
610 615 620

Lys Tyr Cys Val Gly Arg Arg Met Lys Phe Lys Ser Cys Asn Thr Glu
625 630 635 640

Pro Cys Leu Lys Gln Lys Arg Asp Phe Arg Asp Glu Gln Cys Ala His
645 650 655

Phe Asp Gly Lys His Phe Asn Ile Asn Gly Leu Leu Pro Asn Val Arg
660 665 670

Trp Val Pro Gln Tyr Ser Gly Ile Leu Met Lys Asp Arg Cys Lys Leu
675 680 685

Phe Cys Arg Val Ala Gly Asn Thr Ala Tyr Tyr Gln Leu Arg Asp Arg
690 695 700

Val Ile Asp Gly Thr Pro Cys Gly Gln Asp Thr Asn Asp Ile Cys Val
705 710 715 720

Gln Gly Leu Cys Arg Gln Ala Gly Cys Asp His Val Leu Asn Ser Lys
725 730 735

Ala Arg Arg Asp Lys Cys Gly Val Cys Gly Gly Asp Asn Ser Ser Cys
740 745 750

Lys Thr Val Ala Gly Thr Phe Asn Thr Val His Tyr Gly Tyr Asn Thr
755 760 765

Val Val Arg Ile Pro Ala Gly Ala Thr Asn Ile Asp Val Arg Gln His
770 775 780

Ser Phe Ser Gly Gln Thr Asp Asp Asp Asn Tyr Leu Ala [illegible] [illegible] [illegible]
[illegible] 790 795 800

Ser Tyr [illegible] [illegible] Phe Leu [illegible] Asn Gly [illegible] Phe Val Val [illegible] [illegible] [illegible]
[illegible] [illegible] [illegible]

Lys Asn [illegible] Ile Arg Ile Gly Asn Ala Val Val Gln Tyr Ser Gly [illegible]
820 825 830

Glu Thr Ala Val Glu Arg Ile Asn Ser Thr Asp Arg Ile Gln Gln Gln
835 840 845

Leu Leu [illegible] Gln Val Leu Ser Val Gly Lys Leu Tyr Asn Pro Asp Val
850 855 860

Glu Arg Lys Arg Lys Leu Val Cys Thr Arg Glu Ser Asp Gln Leu Thr
900 905 910

Val Ser Asp Gln Arg Cys Asp Arg Leu Pro Gln Pro Gly His Ile Thr
915 920 925

Glu Pro Cys Gly Thr Asp Cys Asp Leu Arg Trp Ala Thr Val Phe Ser
930 935 940

Arg Pro Leu
945

<210> 8

<211> 4086

<212> DNA

<213> human

<400> 8

ggtcgtggtg ctggagttta agttgagtag taggaatgcg gtagtagtta ggataatata
60

aatagttaaa ttaagaatgg ttatgttagg gttgtacggt agaactgcta ttattcatcc
120

tatgtggtta attgagcagt atgctaagat tttgcgtagc tgggtttggt ttaatcaacc
180

ttaactgcct gctatgatcc ataagtttga gagagtgagg agaaggctta cgtttagtga
240

acgagagatt tggtatatga ttgagatggg ggctagtttt tgtcatgtga gaagaagcag
300

tccggatgtc agaggggtgc cttggctaac ctctggcact cagaagtgaa agggggctat
360

tcctagtttt attgctatac ccattatgat tattaatgat gagtattgat tggtagtatt
420

ggttatggtt cattgtccgg agagtatatt gttgaagagg atagctatta gaaggattat
480

ggatgcggtt acttgcgtga ggaaatactt gatggcagct tctgtggaac gagggtttat
540

ttttttggtt agaactggaa taaaagctag catgtttatt tctaggccta ctcaggtaaa
600

aaatcagtgc gagcttagcg ctgtgatgag tgtgcctgca aagatggtag agtagatgac
660

gggggagggg tgggaagcac catgcagttt gtatcctggg ccacactgct aacgctcctg
720

gtgcgggacc tggccgagat ggggagccca gacgccgcgg cggccgtgcg caaggacagg
780
ctgcatccga ggcaagtgaa attattagag accctgagcg aatacgaaat cgtgtctccc
840
atccgagtga acgctctcgg agaacccttt cccacgaacg tccacttcaa aagaacgcga
900
cggagtatta actctgccac tgacccctgg cctgccttcg cctcctcctc ttcctcctct
960
acctcctccc aggcgcatta ccgcctctct gccttcggcc agcagtttct atttaatctc
1020
accgccaatg ccggatttat cgctccactg ttcactgtca ccctcctcgg gacgcccggg
1080
gtgaatcaga ccaagtttta ttccgaagag gaagcggaac tcaagcactg tttctacaaa
1140
ggctatgtca ataccaactc cgagcacacg gccgtcatca gcctctgctc aggaatgctg
1200
gggacatccc ggtctcatga tggggattat tttattgaac cactacagtc tatggatgaa
1260
caagaagatg aagaggaaca aaacaaaccc cacatcattt ataggcgcag cgccccccag
1320
agagagccct caacaggaag gcatgcatgt gacacctcag aacacaaaaa taggcatagt
1380
aaagacaaga agaaaaccag agcaagaaaa tggggacaaa ggattaacct ggctggtgac
1440
[illegible] taaacagcgg cttagcaaca [illegible] [illegible] [illegible]
1500
[illegible]
[illegible]
[illegible]
[illegible]
[illegible] [illegible] [illegible] [illegible] [illegible] [illegible]
[illegible]
ggaattt[illegible] [illegible] tattgtgaac ttaattgtga tttataatg[illegible] [illegible]
1740
[illegible] [illegible]tttaatg[illegible] [illegible] [illegible] [illegible] [illegible]
[illegible]

ccctatagaa gctgttctat tagtgaagat agtggattga gtacagcttt tacgatcgcc
1980
catgagctgg gccatgtgtt taacatgcct catgatgaca acaacaaatg taaagaagaa
2040
ggagttaaga gtccccagca tgtcatggct ccaacactga acttctacac caacccctgg
2100
atgtggtcaa agtgtagtcg aaaatatatc actgagtttt tagacactgg ttatggcgag
2160
tgtttgctta acgaacctga atccagaccc taccctttgc ctgtccaact gccaggcatc
2220
ctttaagacg tgaataaaca atgtgaattg atttttggac caggttctca ggtgtgccca
2280
tatatgatgc agtgcagacg gctctggtgc aataacgtca atggagtaca caaaggctgc
2340
cggacttagc acacaccctg ggccgatggg acggagtgcg agcctggaaa gcactgcaag
2400
tatggatttt gtgttcccaa agaaatggat gtccccgtga cagatggatc ctggggaagt
2460
tggagtccct ttggaacctg ctccacaaca tgtggagggg gcatcaaaac agccattcga
2520
gagtgcaaca gaccagaacc aaaaaatggt ggaaaatact gtgtaggacg tagaatgaaa
2580
tttaagtcct gcaacacgga gccatgtctc aagcagaagc gagacttccg agatgaacag
2640
tgtgctcact ttgacgggaa gcatttaaac atcaacggtc tgcttcccaa tgtgcgctgg
2700
gtccctaaat acagtggaat tctgatgaag gaccggtgca agttgttctg cagagtggca
2760
gggaacacag cctactatca gcttcgagac agagtgatag atggaactcc ttgtggccag
2820
gacacaaatg atatctgtgt ccagggcctt tgccggcaag ctggatgcga tcatgtttta
2880
aactcaaaag cccggagaga taaatgtggg gtttgtggtg gcgataattc ttcatgcaaa
2940
acagtggcag gaacatttaa tacagtacat tatggttaca atactgtggt ccgaattcca
3000
gctggtgcta ccaatattga tgtgcggcag cacagtttct caggggaaac agacgatgac
3060
aactacttag ctttatcaag cagtaaaggt gaattcttgc taaatggaaa ctttgttgtc
3120

acaatggcca aaagggaaat tcgcattggg aatgctgtgg tagagtacag tgggtccgag
3180

actgccgtag aaagaattaa ctcaacagat cgcattgagc aagaactttt gcttcaggtt
3240

ttgtcggtgg gaaagttgta caacccccgat gtacgctatt ctttcaatat tccaattgaa
3300

gataaacctc agcagtttta ctggaacagt catgggccat ggcaagcatg cagtaaaccc
3360

tgccaagggg aacggaaacg aaaacttgtt tgcaccaggg aatctgatca gcttactgtt
3420

tctgatcaaa gatgcgatcg gctgccccag cctggacaca ttactgaacc ctgtggtaca
3480

gactgtgacc tgaggtgggc cactgttttc tcaaggcctt tataaatgaa ttgtgagagt
3540

cttgcaggag gtcccagcag gagaagcaaa aggaggggat gccggtcttt agttcccctt
3600

tcttgtgttt cagtgaaata agctttaacc aattctccat ccctctggaa ctgattatcc
3660

aagacataca tgtgcagatt tcttgttcac ctaagaatta aaaatagcta atagagtatg
3720

gcacttgcca aaaaaaattc agttgatcct cacmacttac tggtaggta ttagcattat
3780

[illegible]
3840

[illegible]
3900

[illegible]

[illegible]

[illegible]
4080

cctatt
4086

<220>
<223> primer

<400> 9
gatcgcggcc gctatggtgg acacgtggcc tctatggctc c
41

<210> 10
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 10
tgaggcttcc agggccgata actgtgcaga gcactcaccc cat
43

Claims

1. An MPTS protein selected from the group consisting of MPTS-15, MPTS-10, MPTS-19 and MPTS-20, wherein said protein is present in other than its natural environment.

2. The protein according to claim 1, wherein said protein has an amino acid sequence substantially identical to the sequence of SEQ ID NO:01, 03, 05 or 07.

3. A nucleic acid present in other than its natural environment, wherein said nucleic acid has a nucleotide sequence encoding an MPTS protein selected from the group consisting of MPTS-15, MPTS-10, MPTS-19 and MPTS-20.

4. A nucleic acid according to claim 3, wherein said nucleic acid has a nucleic acid sequence that is the same as or substantially identical to the nucleotide sequence of SEQ ID NO:02, 04, 06 or 08.

5. An expression cassette comprising a transcriptional initiation region functional in an expression host, a nucleotide sequence according to claims 3 or 4 under the transcriptional regulation of said transcriptional initiation region, and a transcriptional termination region functional in said expression host.

6. A cell comprising an expression cassette according to claim 5 as part of an extrachromosomal element or integrated into the genome of a host cell as a result of introduction of said expression cassette into said host cell.

7. The cellular progeny of the host cell according to claim 6.

8. A monoclonal antibody binding specifically to an MPTS protein according to claim 1

isolating said protein substantially free of other proteins.

10. An MPTS protein as claimed in claim 1 or 2, whenever produced by the process of claim 9.

11. A method of screening to identify MPTS modulatory agents, said method comprising:

contacting an MPTS protein according to claim 1 with a substrate in the presence of an potential modulatory agent; and

determining the effect of said modulatory agent on the activity of said protein.

12. The method according to claim 11, wherein said substrate comprises a glu-ala bond.

13. The method according to claim 12, wherein said substrate is aggrecan or a fragment thereof.

14. A method of treating a host suffering from a disease condition associated with MPTS activity specifically wherein said disease condition is characterized by the presence of aggrecan cleavage products, said method comprising:

administering to said host an MPTS modulatory agent, specifically an antagonist.

15. Use of a MPTS modulatory agent, obtainable or obtained by the method claimed in claim 11 for the preparation of a medicament for the treatment of a disease condition associated with MPTS activity, specifically wherein said disease condition is characterized by the presence of aggrecan cleavage products, like arthritis.

FIG. 1A
MPTS-15: (2879 bp)

GTCTCTGCTGGTGCTTCTCCCAAGACTATCTTGAAGGTGGGCTGTTTGCCTTTCGTGAACA
CATTCTTGGTAT
(SEQ ID NO:02)

FIG. 1B
>ORF(frame +1)

MEILWKTLTWILSLIMASSEFHSDHRLSYSSQEEFLTYLEHYQLTIPIRVDQNGAFLSFTV
KNDKHSPRRRSMDPIDPQQAVSKLFFKLSAYGKHFHLNLTLNTDFVSKHFTVEYWGKDGPQ
WKHDFLDNCHYTGYLQDQRSTTKVALSNCVGLHGVIATEDEEYFIEPLKNTTEDSKHFSYE
NGHPHVIYKKSALQQRHLYDHSHCGVSDFTRSGKPWWLNDTSTVSYSLPINNTHIHHRQKR
SVSIERRVETLVVADFMMVGYHGRKDIEHYILSVMNIVAKLYRDSSLGNVVNIIVARLIVL
TEDQPNIEINHHADKSLDSFCKWQKSILSHQSDGNTIPENGIAHHDNAVLTPYDICTYKN
KPCGTLGLASVAGMCEPERSCSINEDIGLGSAFTIAHEIGHNFGMNHDGIGNSCGTKGHEA
AKLMAAHITANTNPFSWSACSRDYITSFLDSGRGTCLDNEPPKRDFLYPAVAPGQVYDALR
QCRFQYGATSRQCKYGEVCRELWCLSKSNRCVTNSIPAAEGTLCQTGNIEKGWCYQGDCVP
FGTWPQSIDGGWGPWSLWGECSRTCGGGVSSSLRHCDSPAPSGGGKYCLGERKRYRSCNTD
PCPLGSRDFREKQCADFDNMPFRGKYYNWKPYTGGGVKPCALNCLAEGYNFYTERAPAVID
GTQCNADSLDICINGECKHVGCDNILGSDARLDRCRVCGGDGSTCDAIEGFFNDSLPRGGY
MEVVQIPRGSVHIEVREVAMSKNYIALKSEGDDYYINGAWTIDWPRKFDVAGTAFHYKRPT
DEPESLEALGPTSENLIVMVLLQEQNLGIRYKFNVPITRTGSGDNEVGFTWNHQSWSECSA
TCAGGKMPTRQPTQRARWRTKHILSYALCLLKKLIGNISCRFASSCNLPKETLL*LYYIPF
VFNLM*FVQICW*NTSWHNECLCWCFSQDYLEGGLFAFREHILG
(SEQ ID NO (1)

FIG. 1C

Align MP15 with ADAMTS-6(in public data base)

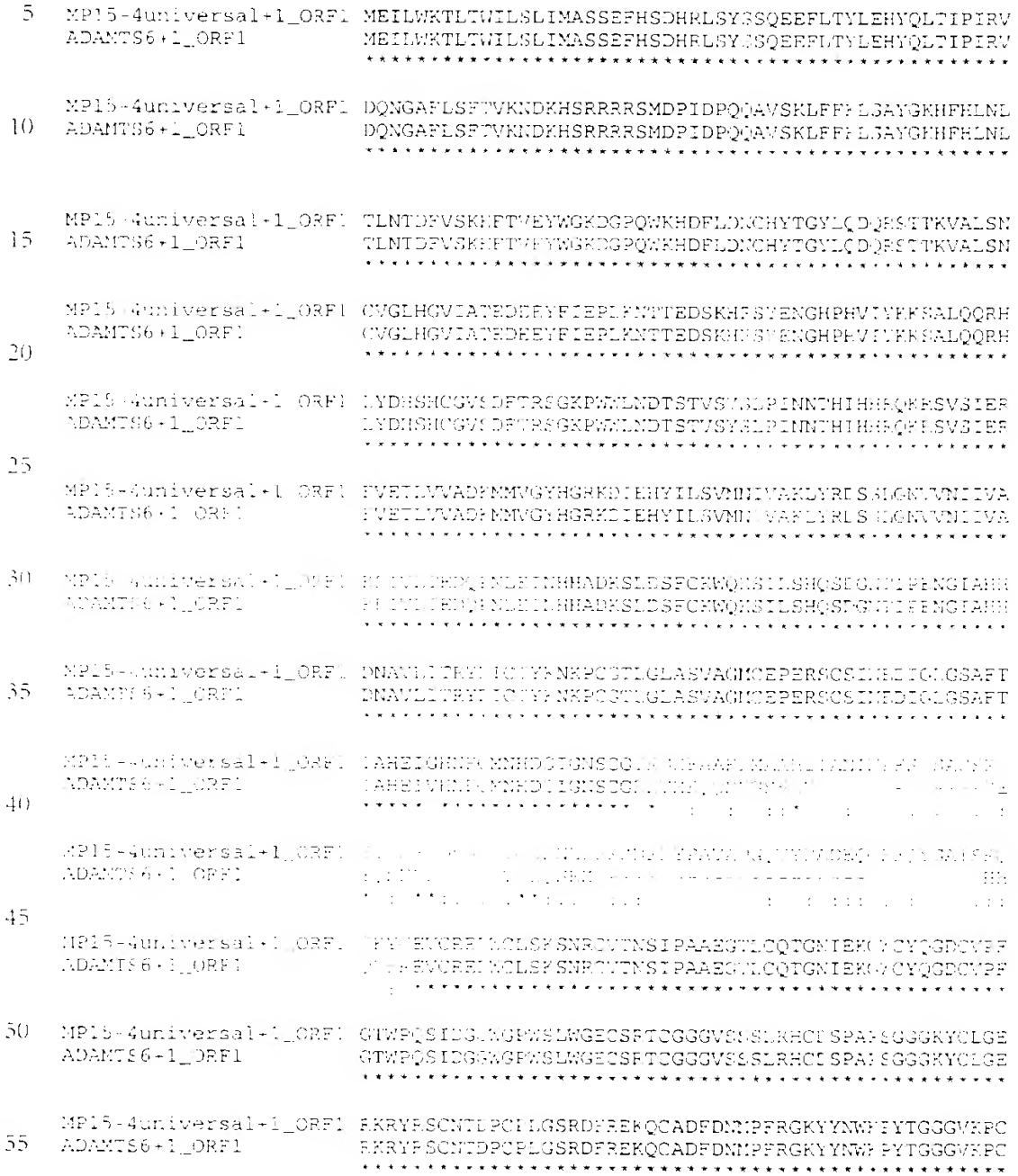


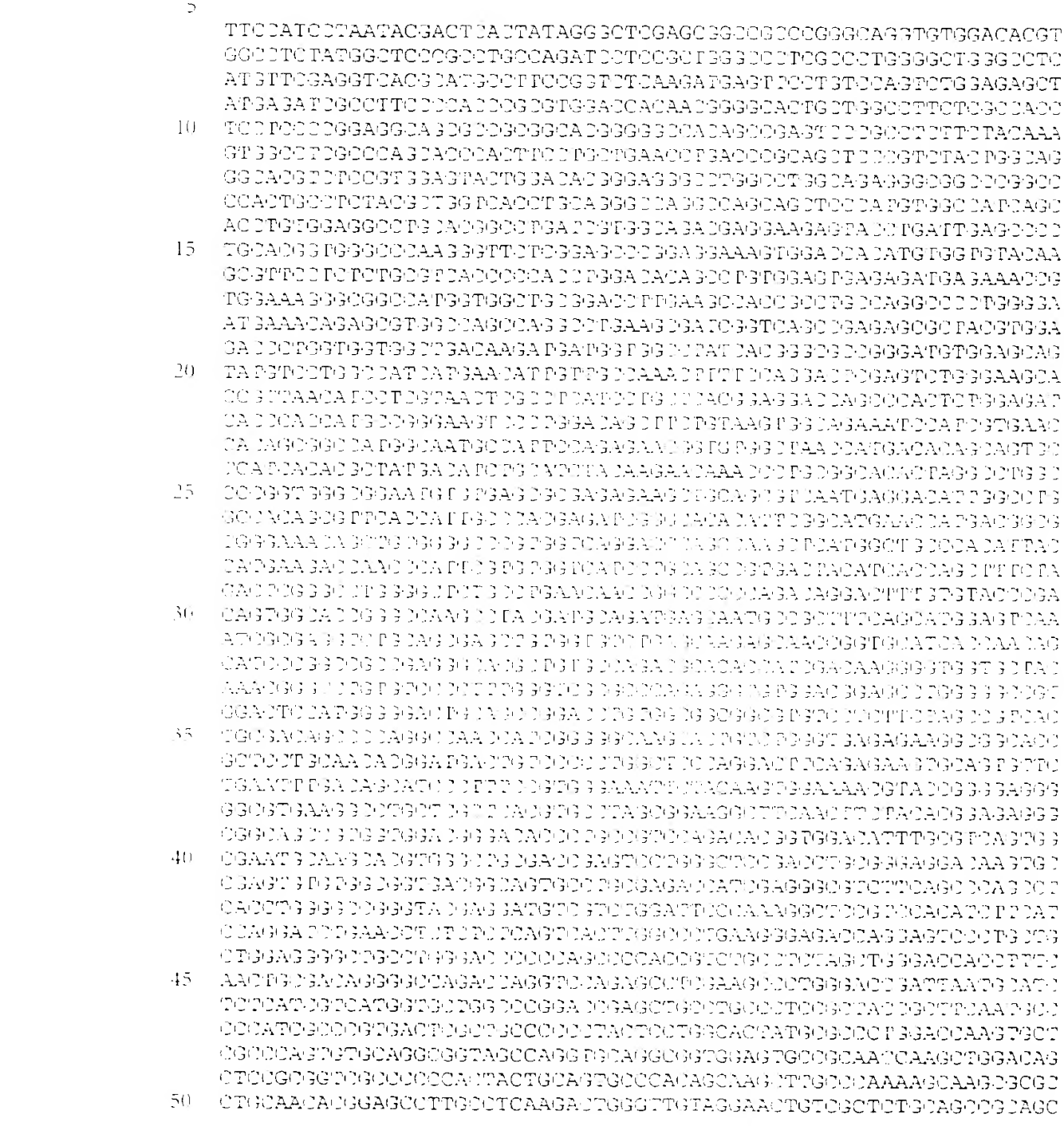
FIG. 1C (Cont)

Align MP15 with ADAMTS-6(in public data base):

```
MP15-4universal+1_ORF1 ALNCLAEGYNFYIERAPAVIDGTQCNADSLDICINGECKHVGCDNILGSD
ADAMTS6+1_ORF1         ALNCLAFGYNFYIERAPAVIDGTQCNADSLDICINGECKHVGCDNILGSD
                       **************************************************

MP15-4universal+1_ORF1 AREDRCRVCGGDGSTCDAIEGFFNDSLPRGGYMEVVQIPRGSVHIEVREV
ADAMTS6+1_ORF1         AREDRCRVCGGGGSTCDAIEGFFNDSLPRGGYMEVVQIPRGSVHIEVREV
                       ********** ***********************************

MP15-4universal+1_ORF1 AMSHNYIALKSEGDDYYINGAWTIDWPRKFDVAGTAFHYKRPTDEPESLE
ADAMTS6+1_ORF1         AMSHNYIALKSEGDDYYINGAWTIDWPRKFDVAGTAFHYKRPTDEPESLE
                       **************************************************

MP15-4universal+1_ORF1 ALGPTSENLIVMILLQEQNLGIRYKFNVPIIRTGSGDNEVGFTWNHQSWS
ADAMTS6+1_ORF1         ALGPTSENLIVMILLQEQNLGIRYKFNVPIIRTGSGLNEVGFTWNHQPWS
                       ************************************************** **

MP15-4universal+1_ORF1 ECSATCAGGKMPTRQPTQRARWRTKHILSYALCLLKKLIGNISCRFASSC
ADAMTS6+1_ORF1         ECSATCAGGKMPTRQPTQRARWRTKHILSYALCLLKKLIGNISCRFASSC
                       **************************************************

MP15-4universal+1_ORF1 NLPKETLL (SEQ ID NO:01)
ADAMTS6+1_ORF1         NLAKETLL (GenBank Accession No. AB140674)
                       ******..
```

FIG. 2A

MP10-full-length



TTGCGATGCAAGGCGTGCGCAGCCGCTCGGTCGTGTGCCAAGCGCCGCGTCTCTGCCGCGA
AGAAAAAGGCGCTGGACGACAGCGCATGCCCGCAGCCGGGCCCACCTGTACTGAGGCCTGCC
ACGGCCCCACTTGCCCTCCGGAGTGGCGGCCCTCGACTGGTCTGAGTGCACCCCCAGCTGC
GGGCCGGGCCTCCGCCACCGCGTGGTCCTTTGCAAGAGCGCAGACCACCGCGCCACGCTGC
CCCCGGCGCACTGCTCACCCGCCGCCAAGCCACCGGCCACCATGCGCTGCAACTTGCGCCG
CTGCCCCCCCGGCCCGCTGGGTGGCTGGCGAGTGGGGTGAGTGCTCTGCACAGTGCGGCGTC
GGGCAGCGGCAGCGCTCGGTGCGCTGCACCAGCCACACGGGCCAGGCGTCGCACGAGTGCA
CGGAGGCCCTGCGGCCGCCGACTAGTAAGCTTCGTCGACCCGGGAATTAATTCCGGACCGG
TACCTGCAGGCGTACCAGCTTTCCCTATAG

(SEQ ID NO: 04)

FIG. 2B

>MP10-full-length+3_ORF1 Translation of MP10-full-length in frame +3, ORF 1, threshold 50

MAPACQILRWALALGLGLMFEVTHAFRSQDEFLSSLESYEIAFPTRVDHNGALLAFSPPPP
ERQRRGTGATAESRLFYKVASPSTHFLLNLTRSSRLLAGHVSVEYWTREGLAWQRAARPHC
LYAGHLQGQASSSHVAISTCGGLHGLIVADEEEYLIEPLHGGPKGSRSPEESGPHVVYKRS
SLRHPHLDTACGVRDEKPWKGRPWWLRTLKPPPAERLGNETERGQPGLKRSVSRERYVETL
VVADKMMVAYHGERDVEQYVLAIMNIVAKLFQDSSLGSTVNILVTRLILLTEDQPTLEITH
HAGKSLDSFCKWQKSIVNHSGHGNAIPENGVANHDTAVLITRYDICIYKNKPCGTLGLAPV
GGMCERERSCSVNEDIGLATAFTIAHEIGHTFGMNHDGVGNSCGAEGQDPAKLMAAHITMK
TNPFVWSSCSRDYITSFLDSGLGLCLNNRPPRQDFVYPTVAPGQAYDADEQCRFQHGVKSR
CLQRAVVSEQEQPVHHQQHPGERGHAVPDAHHRQGVVLQTGLCPLWVAPRGCGFSLGAVDS
MGDCSRTCGGGVSSSSFHCDSPRPTIGGHYCLGERRRHRSCNIDDCPPGSQDFREVQCSEF
DSIPFRGKFYKWRTYRGGGVKACSLTCLAEGFNFYTERAAAVVDGTPCRPDTVDICVSGEC
KHVGCDRVLGSDLREDKCRVCGGDGSACHTIEGVFSPASPGAGYEDVVWIPKGSVHIFIQD
LNLSLSHLALKGDQESLLLEGLPGTPQPHRLPLAGTTFQLRQGPDQVQSLEALGPINASLI
VMVLARTELPALRYRFNAPIARDSLPPYSVHYAPNTKCSPSVQAVARCFRWSAATKLDSSA
VAFHYCRAHSKLAQKQARLQHGALPQDWAVGTVALQPQLAMQGVRSRSVVCQAPRLCREEK
ALDDSACPQPRPEVLRPATAPLALRSGGPRLV*

(SEQ ID NO:03)

FIG. 3A

MP19

ACCGGTGTCCCCTGTGACGAGGCCCAGCAGCCAGCAGCGAAGTCACCTGCTCTCTGCCAC
TCTGTCGGTGGCCCCTGGGCACACTGGGCCCTGAAGGCTCAGGCAGCGGCTCCTCCAGCCA
CGAGCTCTTCAACGAGGCTGACTTCATCCCGCACCACCTGGCCCCACGCCCTTCACCCGCC
TCATCACCCAAGCCAGGCACCATGGGCAACGCCATTGAGGAGGAGGCTCCAGAGCTGGACC
TGCCGGGGCCCGTGTTTGTGGACGACTTCTACTACGACTACAATTTCATCAATTTCCACGA
GGATCTGTCCTACGGGCCCTCTGAGGAGCCCGATCTAGACCTGGCGGGGACAGGGGACCGG
ACACCCCCACCACACAGCCGTCCTGCTGCGCCCTCCACGGGTAGCCCTGTGCCTGCCACAG
AGCCTCCTGCAGCCAAGGAGGAGGGGGTACTGGGACCTTGGTCCCCGAGCCCTTGGCCTAG
CCAGGCCGGCCGCTCCCCACCCCCACCCTCAGAGCAGACCCCTGGGAACCCTTTGATCAAT
TTCCTGCCTGAGGAAGACACCCCCATAGGGGCCCCAGATCTTGGGCTCCCCAGCCTGTCCT
GGCCCAGGGTTTCCACTGATGGCCTGCAGACACCTGCCACCCCTGAGAGCCAAAATGATTT
CCCAGTTGGCAAGGACAGCCAGAGCCAGCTGCCCCCTCCATGGCGGGACAGGACCAATGAG
GTTTTCAAGGATGATGAGGAACCCAAGGGCCGCGGAGCACCCCACCTGCCCCCGAGACCCA
GCTCCACGCTGCCCCCTTTGTCCCCTGTTGGCAGCACCCACTCCTCTCCTAGTCCTGACGT
GGCGGAGCTGTGGACAGGAGGCACAGTGGCCTGGGAGCCAGCTCTGGAGGGTGGCCTGGGG
CCTGTGGACAGTGAACTGTGGCCCACTGTTGGGGTGGCTTCTCTCCTTCCTCCTCCCATAG
CCCCTCTGCCAGAGATGAAGGTCAGGGACAGTTCCCTGGAGCCGGGGACTCCCTCCTTCCC
AGCCCCAGGACCAGGCTCATGGGACCTGCAGACTGTGGCAGTGTGGGGGACCTTCCTCCCC
ACAACCCTGACTGGCCTCGGGCACATGCCTGAGCCTGCCCTGAACCCAGGACCCAAGGGTC
AGCCTGAGTCCCTCAGCCCTGAGGTGCCCCCTGAGCTCTAGGCTGCTGTCCACACCAGCTTG
GGACAGCCCCGGCCAACAGCCACAGAGTCCCCTGAGACCCAGCCGCCTGGCTCCCAGCCCTGGCT
GAAGCGGGGCCCCCCGCGGACCCGTTGGTTGTCAGGAACGCCAGCTGGCAAGCGGGAAACT
GGAGCGAGTGCTCTACCACCTGTGGCCTGGGTGCGGTCTGGAGGCCGGTGCGCTGTAGCTC
CGGCCGGGATGAGGACTGCGCCCCCGCTGGCCGGCCCCAGCCTGCCCGCCGCTGCCACCTG
CGGCCCTGTGCCACCTGGCACTCAGGCAACCTGGAGTAAGTGCCTCCCGCAGCTGCGGCGGAG
GTTCCTCAGTGCGGGACGTGCAGTGTGTGGACACACGGGACCCTCCGGCCACTGCGGCCCCT
CCATTGTCAGCCCGGGCCCTGCCAAGCCGCCTGCGCACCGGCCCCTGCGGGGCCCAGCCCTGC
CTCAGCTGGTACACATCTTCCCTGGAGGGAGTGCTCCGAGGCCCTGTGGCGGTGGTGAGCAGC
AGCGTCTAGTGACCCTGCCCGGAGCCAGGCCCTCTGCGAGGAGGCCGCTGAGACCCAACACCAC
CCGGCCCTGCAACACCCACCCCCTGCACGCAGTGGGTGGTGGGGCCCTGGGGCCAGTGCTCA
GCCCCCTGTGGTGGTGGTGTCCAGCGGCCGCCCTGGTCAAGTGTGTCAACACCCAGACAGGGC
TGCCCGAGGAAGACAGTGACCAGTGTGGCCACGAGGCCCTGGCCCTGAGAGCTCCCGGCCGTG
TGGCACCGAGGATTGTGAGCCCCGTCGAGCCCTCCCCCGCTGTGAGCCGGGACCGCCTGTCCTTC
GGGTTCTGCGAGACGCTGCGCCCTACTGGGCCGCTGCCAGCTGCCCACCATCCGCACCCAGT
GCTGCCGCTCGTGCTCTCCGCCCAGCCACGGCGCCCCCTCCCGACGCCATCAGCGGGTTGC
CCGCCGCTGACTGTGCCAGGATGCACAGACCGACCGACAGACCTCAGTGCCCACCACGGGC
TGTGGCGGAGCTCCCGCCCCCTGCGCCCTAATGGTGCTAACCCCCTCTCACTACCCAGCAG
CAGGCTGGGGACCTCCTCCCCCTCAAAAAAGGTATTTTTTATTTCTAACAGTTTGTGTAAC
ATTTATTATGATTTACATAAATGAGCATCTACCATTCCAAAGCAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAA (SEQ ID NO: 06)

FIG 3B

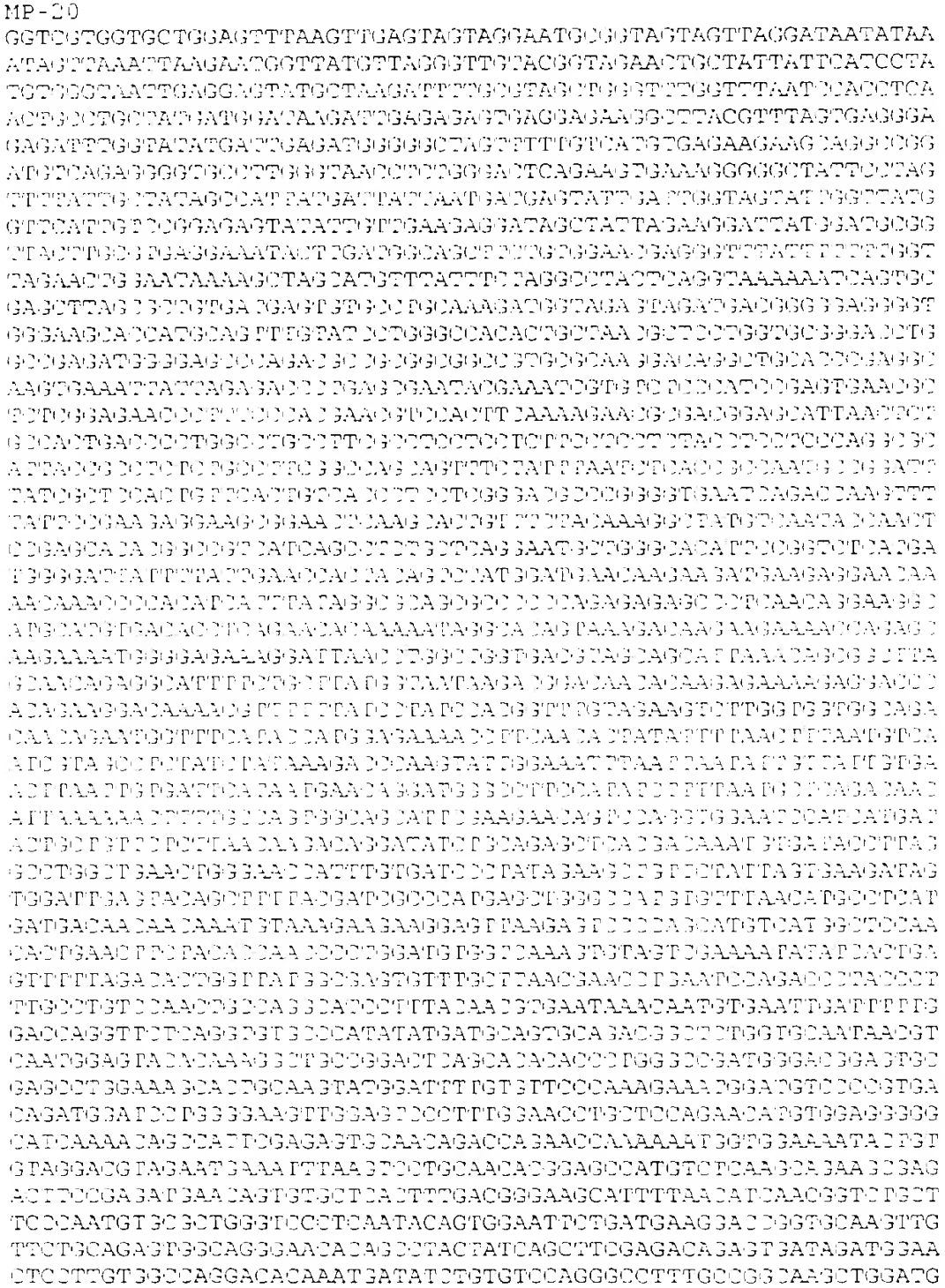
>MP19-full-length+1_ORF1 Translation of MP19-full-length in frame +1, ORF 1, threshold 50

PVPAMPGGPSPRSPAPLLRPLLLLLCALAPGAPGPAPGRATEGRAALDIVHPVRVDAGGSF
LSYELWPRALRKRDVSVRRDAPAFYELQYRGRELRFNLTANQHLLAPGFVSETRRRGGLGP
AHIPAHTPACHLLGEVQDPELEGGLAAISACDGLKGVFQLSNEDYFIEPLDSAPARPGHAQ
PHVVYKPQAPERLAQRGDSSAPSTCGVQVYPELEPRREPWEQRQQWRRPRLRRLHQRSVSK
EKWVETLVVADAKMVEYHGQPQVESYVLTIMNMVAGLFHDPSIGNPIHITIVRLVLLEDEE
EDLKITHHADNTPKSFCKWQKSINMKGDAHPLHHDTAILLTRKDLCATMCRPCETLGLSHV
AGMCQPHRSCSINEDTGLPLAFTVAHELGHSFGIQHDGSGNDCEPVGKRPFIMSPQLLYDA
APLTWSKCSRQYITRFLDRGWGLCLDDPPAKDIIDFPSVPPGVLYDVSHQCRLQYGAYSAF
CEDMDNVCHTLWCSVGTTCHSKLDAAVDGTRCGENKWCLSGECVPVGFRPEAVDGGWSGWS
AWSICSRSCGMGVQSAERQCTQPTPKYKGRYCVGERKRFRLCNLQACPAGRPSFRHVQCSH
FDAMLYYGRLHTWVPVVNDVNPCELHCRPANEYFAEKLRDAVVDGTPCYQVRASRDLCING
ICKIVGCDFEIDSGAMEDECGVCHGNGSTCHTVSGTFEEAEGLGYVDVGLIPAGAREIRIQ
EVAEAANFLALRSEDPEKYFLNGGWTIQWNGDYQVAGTTFTYARRGNWENLTSPGPCTEPV
WIQLLFQESNPGVHYEYTIHREAGGHQEVFPPVFSWHYGPWTKCTVTCGRGVQRQNVYCLE
RQAGPVDEEHCDPLGRPDDQQRKCSEQPCPARWWAGEWQLCSSSCGPGGLSREAVLCIRSV
GSDEQSALEPPACEHLPRPPTETPCNFHVPCPATWAVGNWSQCSVTCGEGTQRRNVLCTND
TGVPCDEAQQPASEVTCSLPLCRWPLGTLGPEGSGSGSSSHELFNEADFIPHHLAPRPSPA
SSPKPGTMGNAIEEEAPELDLPGPVTVDQFYDYNFINFHEDLSYGPSEMPDLDLAGTGDR
TPPPHSFPAAPSTGSPVPATEPPAAKEEGVLGPWSPSPWPSQAGRSPPPPSEQTPGNPGIN
FLPEEDTPIGAPDLGLPSLSWPRVSTDGLQTPATPESQNDFPVGKDSQSQLPPPWRDFTND
VFHDDEEPKGRGAPHLPPRPGSTLPPLSPVDSTHSSPSPDVAELNTDGFVAMEPALEGGLG
PGDSELYPTVGVASLLPPPIAPLPEMKVRDSSLEPGTPSFPAPGPGSWDLQTVAVNGTPLP
TTLPGLCHMPEPALNPGPKSQPESLSPEVPLSSRLLSTPAMDSPANSHRVPETQPLAPSLA
EAGPPAMPLVVRNASWQAGNWSECSTTCGLGAVWRPVRCSSGRDEDCAPAGRPQPARECHL
RPCATWEWGNWSKCSRSCGGGSSVRDVQCVDTRDLPPLNPFHCQPGPAKPPAHRPCGAQPC
LSWVTSEWRECSEACGGGEQQRLVTCPEPGLCREALRPNTTEPCNTHPCTQWVVGPWGQCS
APCGGGVQRRLVKCVNTQTGLPEEDSDQDGHEAWPESSKPCGTEDCEPVEPPPRCERDRLSF
PPCETLFLLGRCQLPTIPFQLCRSCSPPSHFNPSRGHQRWAPP*

(SEQ ID NO:65)

FIG. 4A

MP-20



CGATCATGTTTTAAACTCAAAAGCCCGGAGAGATAAATGTGGGGTTTGTGGTGGCGATAAT
TCTTCATGCAAAACAGTGGCAGGAACATTTAATACAGTACATTATGGTTACAATACTGTGG
TCCGAATTCCAGCTGGTGCTACCAATATTGATGTGCGGCAGCACAGTTTCTCAGGGGAAAC
AGACGATGACAACTACTTAGCTTTATCAAGCAGTAAAGGTGAATTCTTGCTAAATGGAAAC
TTTGTTGTCACAATGGCCAAAAGGGAAATTCGCATTGGGAATGCTGTGGTAGAGTACAGTG
GGTCCGAGACTGCCGTAGAAAGAATTAACTCAACAGATCGCATTGAGCAAGAACTTTTGCT
TCAGGTTTTGTCGGTGGGAAAGTTGTACAACCCCGATGTACGCTATTCTTTCAATATTCCA
ATTGAAGATAAACCTCAGCAGTTTTACTGGAACAGTCATGGGCCATGGCAAGCATGCAGTA
AACCCTGCCAAGGGGAACGGAAACGAAAACTTGTTTGCACCAGGGAATCTGATCAGCTTAC
TGTTTCTGATCAAAGATGCGATCGGCTGCCCCAGCCTGGACACATTACTGAACCCTGTGGT
ACAGACTGTGACCTGAGGTGGGCCACTGTTTTCTCAAGGCCTTTATAAATGAATTGTGAGA
GTCTTGCAGGAGGTCCCAGCAGGAGAAGCAAAAGGAGGGGATGCCGGTCTTTAGTTCCCCT
TTCTTGTGTTTCAGTGAAATAAGCTTTAACCAATTCTCCATCCCTCTGGAACTGATTATCC
AAGACATACATGTGCAGATTTCTTGTTCACCTAAGAATTAAAAATAGCTAATAGAGTATGG
CACTTGCCAAAAAAAAATTCAGTTGATCCTCACMACTTGCTGGGTAGGTATTAGCATTATGA
TTGAGTCACATTGTACGTGAAAACTTGTTTTGAAAGTCAAAAGAAAAGAGGGAGAACCTCA
TCCCTCAAAGTACCCATAATGACCTATATCTACCGAGAGTGTATACCACCCAGTAGAAGAA
CTCCTACACACCTGAAAGTTGCAGTACACTAAGGTAGCGTCATGGAAGAAACAAGAAGAAA
ATGTATATATGGATGTYTGAGATATTCAAACAATTCTGTGTTTAAGAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAGTACTCTGCGTTGTTACCACTGCTTGCCCTATAGTGAGTCGTATT
(SEQ ID NO:08)

FIG. 4B
>MP20-full-length=1_ORF1 Translation of MP20-full-length in frame +1, ORF 1, threshold 50

MQFVSWATLLTLLVRDLAEMGSPDAAAAVFKDRLHPRQVFLLETLSEYEIVSPIRVNALGE
PFPTNVHFKRTRRSINSATDPWPAFASSSSSSTSSQAHYRLSAFGQQFLFNLTANAGFIAP
LFTVTLLGTPGVNQTEFYSEEFAELKHCFYHGYVNTNSEHTAVISLCSGMLGTFRSHDGDY
FIEPLQSMDEQEDEEEQNKPHIIYPRSAPQREPSTGRHACDTSEHKNRHSKDKKKTRARKW
GERINLAGDVAALNSGLATEAFSAYGNKTDNTREKRTHRRTKRFLSYPRFVEVLVVADNFM
VSYHGENLQHYILTLMSIVASIYFDPSIGNLINIVIVNLIVIHNEQDGPSISFNAQTTLKN
FCQWQHSKNSPGGIHHDTAVLLTRQDICRAHDKCDTLGLAELGTICDPYRSCSISEDSGLS
TAFTIAHELGHVFNMPHDDNNKCKEEGVESPQHVMAPTLNFYTNPWMWSKCSREYITEFLD
TGYGECLLNEPESRPYPLPVQLPGILYNVTKQCELIFGPGSQVCPYMMQCRRLWCNNVNGV
HKGCETQHTPWADGTECEPGKHCKYGFCVPKEMDVPVTDGSWGSWSPFGTCSRTCGGGIKT
AIRECNRPEPKNGGKYCVGRRMKFYSCNTEPCLKQKRDFRDEQCAHFDGFHFNIIGLLPNV
RWVPQYSGILMKDRCKLFCRVAGNPAYYQLRDFVIDGTPCGQDTNDICVQGLCRQAGCDHV
LNSKARRDKCGVCGGDNSSCKTVAGTFNTVHYGYNTVVRIPAGATNIDVRQHSFSGETDDD
NYLALSSSKGEFLLNGNFVVTMAKREIRIGNAVVEYSGSETAVERINSTDRIEQELLLQVL
SVGKLYNPDVRYSFNIPIEDKPQQFYWNSHGPWQACSKPCQGERKRKLVCTRESDQLTVSD
QRCDRLPQPGHITEPCGTDCDLRWATVFSRPL*
(SEQ ID NO:07)